The Dissertation Committee for Yining Zeng Certifies that this is the approved version of the following dissertation:

Probing HIV-1 NC-Induced Nucleic Acid Structural Rearrangement by Single-Molecule Fluorescence Spectroscopy

Committee:

Paul F. Barbara, Supervisor

Robert E. Wyatt

David A. Vanden Bout

Keith J. Stevenson

Zhen Yao

Probing HIV-1 NC-Induced Nucleic Acid Structural Rearrangement by Single-Molecule Fluorescence Spectroscopy

by

Yining Zeng, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2007

Dedication

To my parents

Acknowledgements

I would like to express my deepest gratitude to my advisor, Prof. Paul F. Barbara, for providing me such a great chance to study in his supportive lab. Paul has offered me enormous amount of opportunities to learn the start-to-finish of research project. Many times I have been encouraged by Paul to solve problems by myself, and through the processes I have realized to enjoy every challenge as a unique course to expand my knowledge. His deep insights and broad knowledge in science and his perseverance in pursuing the truth have set a high standard for my future career.

I shall also express my gratitude to the people in Barbara group. Christy Landes and Gonzalo Cosa gave me a lot of instruction and guidance in research. Stephan Link and Zhonghua Yu taught me a lot on microscopes and optics. I shall also thank Yeon-Joo Kim, Xiaojing (Jenny) Ma, Yongjing Zhu, who I have worked with in the HIV project. It is my greatest pleasure to be accompanied by the above undoubtedly ingenious people through the past years.

My parents, who always try their best to seek better education for me, have continuously supported me for five years from the other side of the world. I hereby dedicate this dissertation to them.

Probing HIV-1 NC-Induced Nucleic Acid Structural Rearrangement by Single-Molecule Fluorescence Spectroscopy

Publication No._____

Yining Zeng, Ph. D.

The University of Texas at Austin, 2007

Supervisor: Paul F. Barbara

Reverse transcription of HIV-1 genome involves multiple nucleic acids structural rearrangements chaperoned by nucleocapsid protein (NC). One such critical step is that the DNA transactivation response element (TAR) anneals to it complementary RNA region on the genome. It has been challenging to investigate mechanistic details on the annealing process because of the involvement of heterogeneous sets of protein/nucleic acid complexes. Here, we use single-molecule spectroscopy to study the NC induced melting of nucleic acid structure and the annealing activity of different regions along TAR structure. We find that NC induced secondary fluctuations are limited to the terminal stems, and the mechanism for the fluctuations is complex. By employing complementary targeted oligomers, we kinetically trap and investigate stable states of the putative nucleation complexes for the annealing process. This single molecule spectroscopy method directly probes kinetic reversibility and the chaperone role of NC at various stages along the reaction sequence. The new results lead to detailed understanding of NC chaperoned reverse-annealing and the partially annealed

conformational sub-states. Argininamide, because of its specific binding to the loop regions of TAR, was studied on its specific inhibition to strand transfer. The loop-mediated annealing is found to be more potentially inhibited than stem-mediated one.

Table of Contents

List of Tables

List of Figures

List of Schemes

Chapter 1: Introduction and Dissertation Overview

Overview of Human Immunodeficiency Virus (HIV)

Human immunodeficiency virus (HIV) is a retrovirus that causes acquired immunodeficiency syndrome, the so-called AIDS. Patiences of AIDS will usually die in 5 years after suffering from certain damage to their immune systems.[1] According to AIDS epidemic update published by Joint United Nations Programme on HIV and AIDS (UNAIDS) and World Health Organization (WHO), by November 2006, 39.5 million people are living with HIV, among them 2.3 million are children under 15. In 2006 alone, 2.9 million patients died of AIDS and 4.3 million people were newly infected.[2] Undoubtedly, AIDS has been a pandemic that is threatening the world. To control this destructive pandemic and provide better treatment for patients, it is crucial to understand the source of the disease, the HIV virus.

HIV primarily infects vital cells in the human immune system such as helper T cells, specifically $CD4^+$ T cells, caussing loss of human immune response defense mechanism.[3] HIV enters $CD4^+$ T cells through specific interaction between glycoproteins on its surface and the receptors on the target cell, followed by fusion of the viral envelope with the cell membrane. Upon fusion of cell membrances, the HIV RNA and necessary enzymes such as reverse transcriptase, integrase, ribonuclease and protease, etc. are injected into the cell.

There are two types of HIV virus that can infect humans: HIV-1 and HIV-2. HIV-2 is less transmittable and mostly restricted to West Africa region. HIV-1 is more transmittable and virulent, thus receives more attention. As a retrovirus, HIV-1 carries

two indentical copies of single-stranded RNA as its genomic information. Upon infection of a host cell, the HIV-1 viral reverse transcriptase uses human human tRNA^{Lys,3} as a primer to convert the viral RNA genome into complementary double-stranded DNA, cDNA.[4,5] The cDNA is then transported into nucleus and integrated into the host genome DNA by a viral integrase. Integrated with the host genome DNA, the viral DNA is transcribed into viral mRNA and then translated into viral proteins using host cell's translational and transcriptional machinery. Once viral RNA and proteins are made, new HIV-1 virons are assembled at the plasma membrane of the host cell.

The process of reverse transcription step is particularly interesting because it is error-prone and where mutations could occur. It is such mutations that can cause drug resistance. Reverse transcription of the HIV-1 retrovirus involves a series of nucleic acid rearrangements and nucleic acid/nucleic acid annealing steps, illustrated schematically in Figure 1.1 As a start of the reverse transcription, HIV-1 utilizes host cell's tRNA^{Lys,3} as a primer for minus-strand DNA synthesis. The 3' end of tRNA^{Lys,3} is complementary to an 18-nucleotide sequence on the viral RNA template known as the primer binding site (PBS). Prior to initiation of reverse transcription, the tRNA primer must be unwound and annealed to the HIV RNA genome (Figure 1.1, step 1). Then the viral reverse transcriptase uses viral genome as a template to synthesize complementary DNA from the 3' end of tRNA^{Lys,3}. After initiation of minus-strand DNA synthesis by extension of the tRNA primer 3' end (Figure 1.1, step 2), the newly synthesized minus-strand DNA translocates to the 3' end of the RNA genome (step 3). Following completion of minus-strand strong stop DNA ((-)SS DNA) synthesis and initiation of plus-strand DNA synthesis, the plus-strand DNA also undergoes a strand transfer step (not shown). In all the above steps, the HIV-1 nucleocapsid protein (NC) plays a critical role.

Mature HIV-1 nucleocapsid protein (NCp7) is derived by cleavage from the Gag polyprotein and is a small, highly basic protein containing two zinc finger regions.[6] The HIV-1 NC has been shown to perform a diverse set of crucial functions throughout the retroviral life cycle. For example, NC has been shown to melt the nucleic acid hairpin structures, promote placement of the tRNA onto the PBS *in vitro*,[7-9] promote the strand transfer events[10-14] illustrated in Figure 1.1 and inhibit self-priming, etc. Like the primer annealing step, the minus- and plus-strand transfer steps have also been shown to be facilitated by NC.[10-12,15-19] Either NC or its precursor, Gag, is believed to play a role in dimerization of the RNA genome [20-22] and genomic RNA packaging.[23-25]

Structure and Nucleic Acid-Binding Property of NC

The structure of NC has been studied through NMR, CD, and EXAFS investigations.[26-28] These studies have suggested that specific structures of NC may be involved in RNA recognition and binding. Those structural specifications include the basic residues in the linker region between the zinc fingers[27] and hydrophobic patches formed by aromatic residues on the zinc fingers.[26] Recent NMR structures of NC bound to two different stem-loop elements in the RNA genome, SL2 and SL3, have been determined.[29,30] Interaction of NC's zinc fingers with specific RNA bases was observed in both cases, but other binding features, such as the relative orientation of the zinc fingers, were different in the two structures. These results demonstrate that NC may utilize a variety of interactions in its recognition of and binding to NA structures. Studies of NC's preferences for specific nucleic acid sequences have further shown that, in addition to demonstrating a preference for single-stranded sequences, NC has an affinity for G- [31] and GT-rich sequences[32] (although NC is typically characterized as non-specific binder).
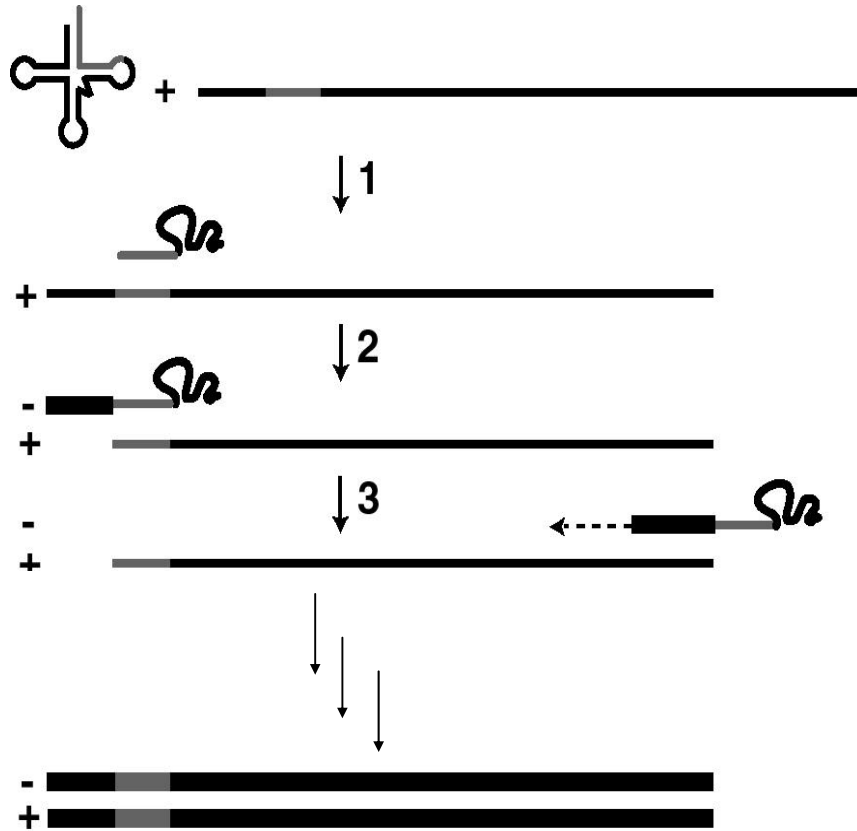
Figure 1.1: Illustration of HIV-1 reverse transcription steps. Step 1: tRNA^{Lys,3} is unwound, and its 3' end 18 complementary nucleotides (shown in gray) are annealed to the primer binding site (PBS) on the RNA genome. Step 2: Minus-strand DNA (thick black line) synthesis is initiated by extension of the tRNA primer. Step 3: The minus-strand strong stop DNA is translocated to the 3' end of the RNA genome in the minus-strand transfer step. After several more steps, including a plus-strand transfer, a full-length proviral double-stranded DNA is synthesized.

NC Induces Partial Melting of Nucleic Acid Structure

As a nucleic acid chaperone, NC catalyzes nucleic acid rearrangements that ultimately result in annealed structures containing the maximum number of base pairs.[6,33] However, the mechanism by which NC destabilizes structured nucleic acid elements and promotes their annealing to other sequences is not well understood. For example, in the case of the tRNA^{Lys,3} priming of DNA synthesis, tRNA^{Lys,3} must be substantially unwound prior to annealing to the PBS of the RNA genome. Evidence for NC-induced unwinding of tRNA has been gathered through CD studies,[34] although other researchers claim that the CD results are more consistent with local base unstacking rather than tRNA unwinding.[35] In collaboration with Prof. Karin Musier-Forsyth, we have employed time-resolved fluorescence resonance energy transfer (FRET) experiments performed to demonstrate that NC does not induce unwinding of the acceptor stem in the absence of the RNA genome acceptor but may promote more subtle changes in tRNA structure.[36] NMR results have further supported the idea that NC does not induce the melting of specific base pairs or the global unwinding of helical domains in the tRNA.[37] Two regions of the tRNA sensitive to catalysis by NC were, however, identified in the same NMR study by relaxation time and exchange measurements.[37] Results from the lab of our collaborator, Prof. Musier-Forsyth, also shows that NC induces more subtle conformational changes in the tRNA tertiary core and A-form helices in the absence of the RNA genome.[38] Collectively, these results indicate that, in the absence of the RNA genome acceptor, NC not only induces the global tRNA unwinding necessary for annealing but also promotes more subtle and transient rearrangements.

However, the conclusions from all of these previous studies are based on ensemble-averaged results from bulk experiments. Due to its intrinsic averaging effect, ensemble/bulk measurements are difficult to interpret and often yield ambiguous results

due to, for example, conformational and kinetic heterogeneity. These are well-known complications in the interpretation of FRET data that are manifested, for example, by multi-exponential fluorescence transients.[39]

NC Promotes Strand Transfer

The minus- and plus-strand transfer steps (Figure 1.1) of HIV reverse transcription raise additional questions regarding the mechanism of NC's chaperone functions. The minus-strand transfer step that follows (-) SS DNA synthesis requires destabilization of a highly structured TAR (Trans-Activation Responsive region) element present in both the RNA genome and in the complementary (-)SS DNA (TAR DNA). Ninty-eight base pairs must be formed during the annealing step of minus-strand transfer *in vivo*, making this a more complex system than tRNA:PBS annealing (18 base pairs). The TAR DNA does not easily undergo minus-strand transfer in the absence of NC [12] due, in part, to TAR-induced self-priming.[19] In the presence of NC, minus-strand transfer occurs at a 3000-fold higher rate than in its absence.[12] Kinetic studies of this step suggest that it is a first-order process,[12] which is unusual for an annealing reaction of this complexity involving such a large number of base pairs. Accordingly, it has been hypothesized that conformational change rather than annealing is the rate-limiting step (RLS) in this case. (NC also promotes the annealing step of the plus-strand transfer although to a lesser extent than in minus-strand transfer.[15]) In addition to the intrinsic interest in the strand transfer processes, these examples are also promising prototypes for the many known complex RNA/DNA annealing "reactions," some of which are considerably more complex, involving multiple co-factors and ATP. For such systems, molecular-level details of the reactions are poorly understood.
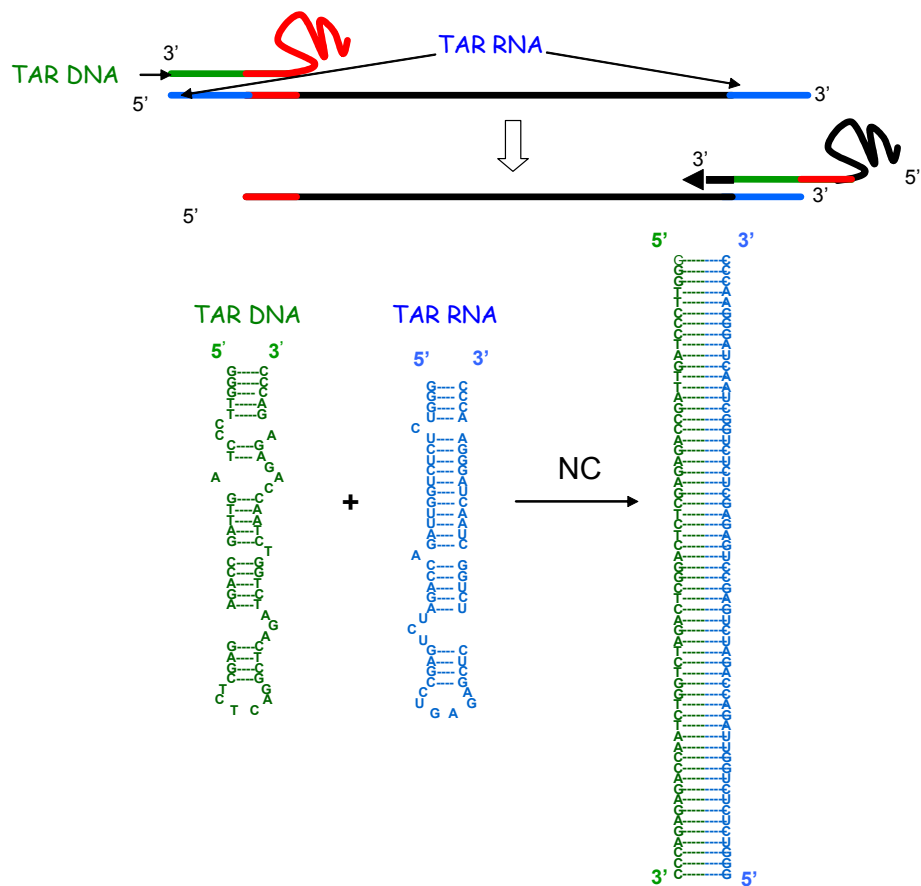
Figure 1.2 Illustration of HIV-1 minus-strand transfer steps. The TAR DNA
attached to the TAR RNA at the 5' end will have to transfer to the 3' end and
anneals to the TAR RNA sequence wherein.

Single-Molecule Spectroscopy and its Application to Studies of NC's Function

The greatest strength of single molecule spectroscopy (SMS) [40-47] lies in its ability to avoid the obscuring effect of heterogeneity found in conventional ensemble measurements of molecular dynamics and chemical kinetics in highly complex systems. SMS methods have been particularly successful in the investigation of biochemical systems including nucleic acids,[48-52] membranes,[53-58] enzymes,[59-61] motor proteins,[62-65] and other proteins and peptides.[66-68] In addition, new single molecule spectroscopic tools and sample preparation techniques have been developed for SMS studies of biological systems. These studies have investigated molecular properties that would be obscured in bulk experiments and, in some cases, have revealed previously unknown pathways, intermediates, or conformations. SMS is especially promising as a mechanistic tool in the investigation of complex biochemical mechanisms with parallel kinetic pathways, multiple transition states, and complex chemical intermediates with poorly defined structures.

The intense single-molecule spectroscopy studies carried out in our group have uniquely well demonstrated that the NC induces partial melting of the TAR DNA. NC/TAR DNA complex is an equilibrium mixture of closed (C) and partially open (Y) TAR DNA/NC forms.[69-74] Single-molecule study of the C/Y equilibration dynamics reveals a strong dependence of the FRET dynamics on both the NC and Mg concentrations, leading to several new insights on the C/Y interconversion mechanism. The C/Y dynamics were also observed to be significantly heterogeneous, indicating that the C/Y inter-conversion is not a simple two-state equilibration but rather must involve multiple conformational intermediates and/or multiple reaction pathways.[69,70] Furthermore, our single-molecule studies also show that NC not only chaperones the annealing of two complementary nucleic acids, but also chaperones the dissociation, or

the reverse annealing of the adduct.[71]   Additionally, single-molecule studies also show that the chaperone ability of NC is also inhibited by argininamide, a small molecule nucleic acid binder.[74]

Overview of Dissertation

In the next chapters, this dissertation will continue to discuss the following topics:

Chapter 2 **.** In this chapter, we investigate the ability of NC to melt the nucleic acid strands. We use time resolved single-molecule fluorescence resonance energy transfer to study NC chaperoned opening of various DNA hairpins over a broader range of conditions. We have found that NC induces a complex mechanism for secondary structure fluctuations with dynamic processes occurring over a wide time range, i.e. ~5 to >250 milliseconds and with the involvement of long-lived intermediates. The dynamic role of DNA loop regions and NC binding/dissociation events are discussed.

Chapter 3. Herein, we investigate the NC-assisted *in vitro* annealing mechanism by a novel multi-step SMS kinetic approach. We immobilize TAR DNA hairpin on surface, then expose hairpin to a multi-step programmed concentration sequence of NC, model complementary targeted-oligonucleotides, and buffer only solutions to initiate the reaction. This controllably sequence SMS method directly probes kinetic reversibility and the chaperone (catalytic) role of NC at various stages along the reaction sequence, i.e. reactants, intermediates, and products. This unique approach provides access to previously inaccessible kinetic processes and rate constants, leading to a more complete and detailed understanding of the ability of NC to promote nucleic acid/nucleic acid rearrangement processes. We also provides the first information on the ability of NC to chaperone "reverse annealing" in single strand transfer and the first observation of partially annealed, conformational sub-states in the annealing mechanism.

Chapter 4. We explore how argininamide modulates the ability of NC to chaperone the annealing reaction as well as to change the secondary structure of TAR DNA. Our studies show that the argininamide inhibitory mechanism involves a shift of the secondary structure of TAR, away from the NC-induced partially open form, "Y" form, an intermediate in reverse transcription, and toward the free closed "C" form. In addition, more potential inhibition of the loop-mediated annealing pathway than stem-mediated annealing is observed. Collectively, these data suggest a molecular mechanism wherein argininamide inhibits NC-facilitated TAR RNA/DNA annealing *in vitro* by interfering with the formation of key annealing intermediates.

Chapter 5. This chapter provides more detailed information on instrumentation and programming for the single-molecule spectroscopy tools used in this dissertation. First, the basic one-color excitation confocal setup for single-molecule fluorescence trajectory measurement and correlation spectroscopy is briefly described. Then a user developed Matlab program for the single molecule trajectory analysis, called SMS developed for Chapter 2 and 4 is introduced. Following that, a more advanced two-color alternative excitation scanning confocal microscope, used in Chapter 3 is introduced. The Labview virtue instruments (VIs) for controlling the microscope are explained in technical details. Finally, the evaluation with respect to signal-to-background and photobleach etc. of a TIRF wide-field microscope is presented. This shows why the wide-field microscope is incapable for measuring slow annealing kinetics but certainly sufficient for dye-aggregating system.

Enjoy reading!

References

(1) Schneider MF, Gange, S. J., Williams, C. M., Anastos, K., Greenblatt, R. M., Kingsley, L., Detels, R., and Munoz, A. (2005) *AIDS* 19: 2009–2018.

(2) *Global AIDS epidemic continues to grow* (Geneva, 21 November 2006) *World Health Organization*

(3) Luciw PA, Potter, S.J., Steimer, K., Dina, D. and Levy, J.A. (1984) *Nature* 312: 760-763.

(4) Kleiman L, Cen S (2004) *Int. J. Biochem. Cell Biol.* 36: 1776-86.

(5) Kleiman L, Halwani R, Javanbakht H (2004) *Curr. HIV Res.* 2: 163-175.

(6) Rein A, Henderson LE, Levin JG (1998) *Trends Biochem Sci* 23: 297-301.

(7) Li X, Quan Y, Arts EJ, Li Z, Preston BD, de Rocquigny H, Roques BP, Darlix J-L, Kleiman L, Parniak MA, Wainberg MA (1996) *J Virol* 70: 4996-5004.

(8) de Rocquigny H, Gabus C, Vincent A, Fournié-Zaluski M-C, Roques B, Darlix J-L (1992) *Proc Natl Acad Sci USA* 89: 6472-6476.

(9) Barat C, Lullien V, Schatz O, Keith G, Nugeyre MT, Grüninger-Leitch F, Barrè-Sinoussi F, LeGrice SFJ, Darlix JL (1989) *Eur J Biochem* 8: 3279-3285.

(10) Peliska JA, Balasubramanian S, Giedroc D, Benkovic SJ (1994) *Biochemistry* 33: 13817-13823.

(11) Rodriguez-Rodriguez L, Tsuchihashi Z, Fuentes GM, Bambara RA, Fay PJ (1995) *J of Biol Chem* 270: 15005-15011.

(12) You JC, McHenry CS (1994) *J Biol Chem* 269: 31491-31495.

(13)   Guo J, Wu T, Anderson J, Kane BF, Johnson DG, Gorelick RJ, Henderson LE, Levin JG (2000) *J Virol* 74: 8980-8988.

(14)   Johnson PE, Turner RB, Wu ZR, Hairston L, Guo J, Levin JG, Summers MF (2000) *Biochemistry* 39: 9084-9091.

(15)   Wu T, Guo J, Bess J, Henderson LE, Levin JG (1999) *J. Virol.* 73: 4794-4805.

(16)   Allain B, Lapadat-Tapolsky M, Berlioz C, Darlix J-L (1994) *EMBO J.* 13: 973-981.

(17)   Brule F, Bec G, Keith G, Grice SFJL, Roques BP, Ehresmann B, Ehresmann C, Marquet R (2000) *Nucleic Acids Res.* 28: 634-640.

(18)   Darlix J-L, Vincent A, Garus C, De Rocquigny H, Roques B (1993) *Genetics* 316: 763-771.

(19)   Guo J, Henderson LE, Bess J, Kane BF, Levin JG (1997) *J. Virol.* 71: 5178-5188.

(20)   Feng Y-X, Copeland TD, Henderson LE, Gorelick RJ, Bosche WJ, Levin JG, Rein A (1996) *Proc. Natl. Acad. Sci. USA* 93: 7577-7581.

(21)   Darlix J-L, Gabus C, Nugeyre M-T, Clavel F, Barré-Sinoussi F (1990) *J Mol Biol* 216: 689-699.

(22)   Sakaguchi K, Zambrano N, Baldwin ET, Shapiro BA, Erickson JW, Omichinski JG, Clore GM, Gronenborn AM (1993) *Proc Natl Acad Sci USA* 90: 5219-5223.

(23)   Gorelick RJ, Gagliardi TD, Bosche WJ, Wiltrout TA, Coren LV, Chabot DJ, Lifson JD, Henderson LE, Arthur LO (1999) *Virology* 256: 92-104.

(24)   Zhang Y, Barklis E (1995) *J Virol* 69: 5716-5722.

(25)   Berkowitz RD, Ohagen Å, Höglund S, Goff S (1995) *J Virol* 69: 6445-6456.

(26) Summers MF, Henderson LE, Chance MR, Bess JW, Jr., South TL, Blake PR, Sagi I, Perez-Alvarado G, Sowder RC, III, Hare DR, Arthur LO (1992) *Protein Sci* 1: 563-574.

(27) Omichinski JG, Clore GM, Sakaguchi K, Appella E, Gronenborn AM (1991) *FEBS Lett* 292: 25-30.

(28) Morellet N, Jullian N, de Rocquigny H, Maigret B, Darlix J-L, Roques BP (1992) *EMBO J* 11: 3059-3065.

(29) de Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF (1998) *Science* 279: 384-388.

(30) Amarasinghe GK, de Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF (2000) *J Mol Biol* 301: 491-511.

(31) Vuilleumier C, Bombara E, Morellet N, Gérard D, Roques BP, Mély Y (1999) *Biochem* 38: 16816-16825.

(32) Fisher RJ, Rein A, Fivash M, Urbaneja MA, Casas-Finet JR, Medaglia M, Henderson LE (1998) *J Virol* 72: 1902-1909.

(33) Tsuchihashi Z, Brown PO (1994) *J Virol* 68: 5863-5870.

(34) Khan R, Giedroc DP (1992) *J Biol Chem* 267: 6689-6695.

(35) Gregoire CJ, Gautheret D, Loret EP (1997) *J Biol Chem* 272: 25143-25148.

(36) Chan B, Weidemaier K, Yip W-T, Barbara PF, Musier-Forsyth K (1999) *Proc Natl Acad Sci USA* 96: 459-464.

(37) Tisné C, Roques BP, Dardel F (2001) *J Mol Biol* 306: 443-454.

(38) Hargittai MRS, Mangla AT, Gorelick RJ, Musier-Forsyth K (2001) *J. Mol. Biol* 312: 985-987.

(39)   Jia Y, Sytnik A, Li L, Vladimirov S, Cooperman BS, Hochstrasser RM (1997) *Proc. Natl. Acad. Sci. USA* 94: 7932-7936.

(40)   Ha T (2001) *Curr. Opin. Struct. Biol.* 11: 287-292.

(41)   Deniz AA, Laurence TA, Dahan M, Chemla DS, Schultz PG, Weiss S (2001) *Annu. Rev. Phys. Chem.* 52: 233-253.

(42)   Nie S, Zare R (1997) *Annu. Rev. Biophys. Biomol. Struct.* 26: 567-596.

(43)   Ambrose WP, Goodwin PM, Jett JH, Van Orden A, Werner JH, Keller RA (1999) *Chem. Rev.* 99: 2929-2956.

(44)   Xie XS, Trautman JK (1998) *Annu. Rev. Phys. Chem.* 49: 441-480.

(45)   Weiss S (2000) *Nature Struct. Biol.* 7: 724-729.

(46)   Moerner WE, Orrit M (1999) *Science* 283: 1670-1676.

(47)   Weiss S (1999) *Science* 283: 1676-1683.

(48)   Grunwell JR, Glass JL, Lacoste TD, Deniz AA, Chemla DS, Schultz PG (2001) *J. Am. Chem. Soc.* 123: 4295-4303.

(49)   Ha T, Zhuang XW, Kim HD (1999) *Proc. Natl. Acad. Sci. USA* 96: 9077-9082.

(50)   Deniz AA, Dahan M, Grunwell JR, Ha TJ, Faulhaber AE, Chemla DS, Weiss S, Schultz PG (1999) *Proc. Natl. Acad. Sci. USA* 96: 3670-3675.

(51)   Eggeling C, Fries JR, Brand L, Gunther R, Seidel CAM (1998) *Proc. Natl. Acad. Sci. USA* 95: 1556-1561.

(52)   Kim HD, Nienhaus GU, Ha T, Orr JW, Williamson JR, Chu S (2002) *Proc. Natl. Acad. Sci.* 99: 4284-4289.

(53)   Ke PC, Naumann CA (2001) *Langmuir*: in press.

(54)   Burden DL, Kasianowicz JJ (2000) *J. Phys. Chem. B* 104: 6103-6107.

(55)   Schutz GJ, Kada G, Pasturshenko VP, Schindler H (2000) *EMBO J.* 19: 892-901.

(56)   Talley CE, Dunn RC (1999) *J. Phys. Chem. B* 103: 10214-10220.

(57)   Schwille P, Korlach J, Webb WW (1999) *Cytometry* 36: 176-182.

(58)   Schmidt T, Schultz GJ, Baumgartner W, Gruber HJ, Schindler H (1996) *Proc. Natl. Acad. Sci. USA* 93: 2926-2929.

(59)   Ha TJ, Ting AY, Liang J, Caldwell WB, Deniz AA, Chemla DS, Schultz PG, Weiss S (1999) *Proc. Natl. Acad. Sci. USA* 96: 893-898.

(60)   Zhuang XW, Bartley LE, Babcock HP, Russell R, Ha TJ, Herschlag D, Chu S (2000) *Science* 288: 2048-2051.

(61)   Lu PH, Xun L, Xie XS (1998) *Science* 282: 1877.

(62)   Mehta AD, Rief M, Spudich JA, Smith DA, Simmons RM (1999) *Science* 283: 1689-1695.

(63)   Sase I, Miyata H, Ishiwata S, Kinosita K (1997) *Proc. Natl. Acad. Sci.* 94: 5646-5650.

(64)   Funatsu T, Harada Y, Tokunaga M, Saito K, Yanagida T (1995) *Nature* 374: 555-559.

(65)   Tokunaga M, Kitamura K, Saito K, Iwane AH, Yanagida T (1997) *Biochem. Biophys. Res. Comm.* 235: 47-53.

(66)   Talaga DS, Lau WL, Roder H, Tang J, Jia Y, DeGrado WF, Hochstrasser RM (2000) *Proc. Natl. Acad. Sci. USA* 97: 13021-13026.

(67)   Zhuang X, Ha T, Kim HD, Centner T, Labeit S, Chu S (2000) *Proc. Natl. Acad. Sci.* 97: 14241-14244.

(68)   Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Schultz PG, Weiss S (2000) *Proc. Natl. Acad. Sci. USA* 97: 5179-5184.

(69)   Cosa G, Harbron EJ, Zeng Y, Liu H-W, O'Connor DB, Eta-Hosokawa C, Musier-Forsyth K, Barbara PF (2004) *Biophysical Journal* 87: 2759-2767.

(70)   Cosa G, Zeng Y, Liu H-W, Landes CF, Makarov DE, Musier-Forsyth K, Barbara PF (2006) *J. Phys. Chem. B* 110: 2419-2426.

(71)   Zeng Y, Liu H-W, Landes CF, Kim YJ, Ma X, Zhu Y, Musier-Forsyth K, Barbara PF (2007) *Proceedings of the National Academy of Sciences* 104: 12651-12656.

(72)   Liu H-W, Cosa G, Landes CF, Zeng Y, Mullen DG, Barany G, Musier-Forsyth K, Barbara PF (2005) *Biophysical Journal* 89: 3470-3479.

(73)   Liu H-W, Zeng Y, Landes CF, Kim YJ, Zhu Y, Ma X, Vo M-N, Musier-Forsyth K, Barbara PF (2007) *Proc Natl Acad Sci USA* 14: 5261-5267.

(74)   Landes CF, Zeng Y, Liu H-W, Musier-Forsyth K, Barbara PF (2007) *J. Am. Chem. Soc.* 129: 10181-10188.

Chapter 2: Evidence for Non-Two-State Kinetics in the Nucleocapsid Protein Chaperoned Opening of DNA Hairpins

Introduction

Nucleic acid binding proteins play a central role in many critical processes in genetics.[1-3] Specific examples include replication of DNA,[4] transcription of DNA to RNA,[5,6] and reverse transcription of RNA to DNA.[7] A well known example is the *E. coli* single-stranded DNA binding protein, SSB, which has been observed to support replication by forming a complex with DNA in which double-stranded DNA is unwound.[4] The present paper is focused on another important nucleic acid binding protein, namely the nucleocapsid protein, NC, of the retrovirus HIV-1.[8] In addition to stabilizing the enveloped virion,[9,10] NC also chaperones the formation of stable nucleic acid duplexes (both RNA and DNA) in several distinct steps in the reverse transcription mechanism of the retrovirus.[11-17] NC destabilizes the secondary structure of the bound nucleic acid in order to promote subsequent annealing steps.[18-22] Investigating the structure and dynamics of NC/nucleic acid complexes is a critical first step for building a molecular level understanding of HIV-1 reverse transcription.

Time-resolved single molecule spectroscopy, SMS, has been applied to study nucleic acid conformational dynamics.[23,24] SMS can be used to unravel the dynamics of bimolecular processes that involve the interconversion of multiple structures. These studies have shown that the secondary and tertiary structures of protein/nucleic acid complexes can be highly dynamic, involving a complex distribution of nucleic acid

structures that inter-convert over a broad distribution of time scales. For example, the complex of reverse transcriptase with DNA was observed to adopt three distinct structures.[25] Single-molecule fluorescence resonance energy transfer, SMFRET, revealed a distribution of rates for the unwinding of DNA by *E. coli* helicase.[26,27] Such complex processes would be difficult to characterize accurately by traditional ensemble methods alone.

RNA folding/unfolding represents a further example of dynamic complexity that can be revealed by SMS. SMFRET was used to examine heterogeneous structural dynamics in *B. subtilis* RNase P RNA.[28] Other single molecule RNA folding studies demonstrated that conformational transformations exhibit dynamic heterogeneity,[29,30] and can occur on timescales that vary by several orders of magnitude.[24,31] Single molecule force measurements have been used to quantify the free energy landscape for nucleic acid folding, and similar complexity has been observed.[32,33] It has been postulated that a strong correlation between structural dynamics and heterogeneous reaction pathways is ubiquitous.[31] Thus, it remains important to characterize the nucleic acid structural dynamics induced by NC.

Recently, we used SMFRET to study biotin-immobilized NC/TAR DNA complexes.[20] TAR DNA is a hairpin structure with complementarity to the transactivation response element, TAR RNA, and is also a component of minus-strand strong-stop DNA.[20] In the minus-strand transfer step of HIV-1 reverse transcription, TAR DNA must anneal to the complementary TAR RNA hairpin located at the 3' end of the viral RNA in order for minus-strand synthesis to continue.[15] SMFRET

measurements of the 3'/5' end-to-end distance fluctuations of TAR DNA *in the absence of NC* revealed that TAR-DNA is in its closed form with a maximum number of Watson-Crick base pairs formed.[20] In contrast, SMFRET studies *in the presence of NC* demonstrated that NC destabilizes the secondary structure of the terminal two loop regions of the initially closed TAR DNA.

Cosa *et al.* proposed[20] that the NC/TAR DNA complex is an equilibrium mixture of closed and partially open TAR DNA/NC forms, denoted in Scheme 2.1 as the **C** and **Y** forms, respectively, where the loop regions are labeled by L1-L4. It is important to note that the **Y** terminology implies only that the Watson-Crick base pairs in the L1-L2 region have been broken due to bound NC, and does not indicate a specific secondary or tertiary structure. Indeed, considering the flexibility of the single-stranded regions and the likelihood of a broad distribution of NC binding-sites, it seems likely that the **Y** form should in fact be structurally and perhaps dynamically heterogeneous.

Previous SMFRET experiments on TAR DNA mutants (see Scheme 2.2) helped assign the observed NC-induced secondary structure fluctuations to the L1-L2 region of TAR. The experiments were made at sufficiently high NC concentration to ensure saturated NC binding, i.e., 1 NC molecule/7 nucleotides (see ref.[8] for review). Under these conditions the observed SMFRET dynamics were qualitatively consistent with a simple kinetically homogeneous **C**/**Y** inter-conversion on the 3-10 ms time scale depending on Mg^{2+} concentration.[34] However, Cosa *et al.* also observed that a small fraction of the NC/DNA complexes exhibited much slower dynamics, which were

ascribed tentatively to NC/DNA complexes that were perturbed by the immobilization process.[20]

The current report is a more extensive and in depth characterization of the secondary structure fluctuations of the L1-L2 region of TAR DNA in the presence of NC. NC-induced dynamics in the L1-L2 region were measured over a larger range of NC, Mg^{2+} and Na^+ concentrations than in previous studies. These studies reveal multiple timescales for 3'/5' end-to-end distance fluctuations, ranging from single milliseconds to >250 ms. The current SMFRET results suggest that the **C/Y** interconversion is more dynamically complex than previously thought, and in fact involves multiple intermediate states and potentially multiple pathways. This is the first report of dynamic complexity for the NC/TAR DNA system, and may have implications for the multiple annealing pathways that have been observed in separate TAR DNA/oligonucleotide binding studies.[35]

5' 3'
L1
L2
NC
L3
L3
L4
L4
C form TAR DNA
Y form TAR DNA

5'GGGT T C CCTA   CAGA GA G A C C C 3'
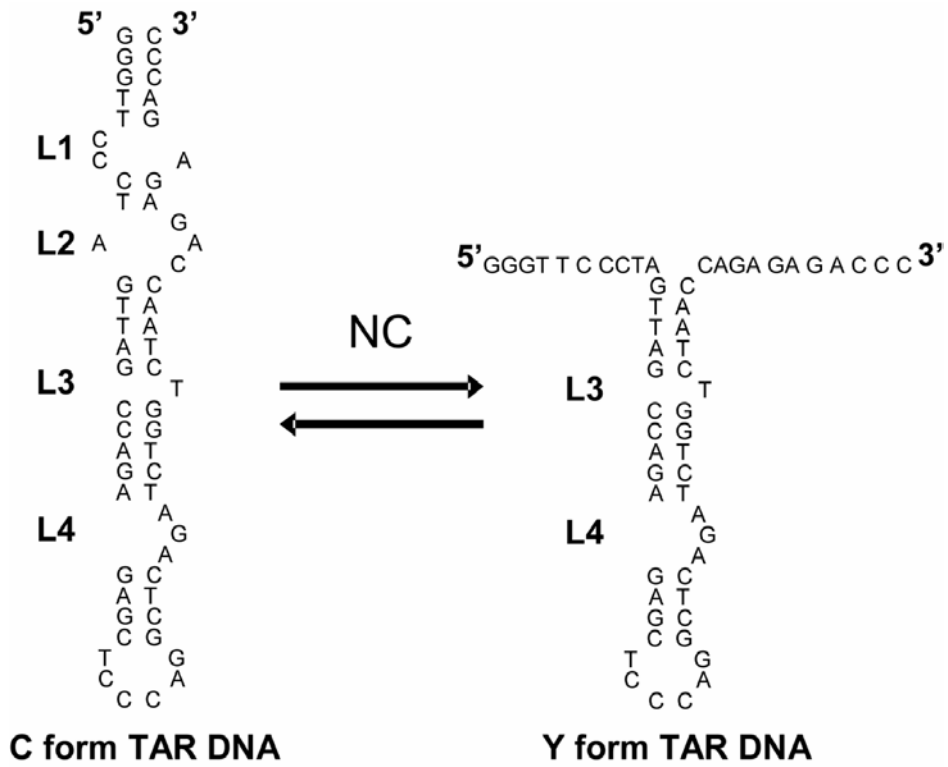
Scheme 2.1 (left) The primary and secondary structure of TAR DNA is shown in its closed, **C**, form with the four loop regions labeled. (right) In the presence of NC, the secondary structure in the L1-L2 regions is disrupted, termed the **Y** form.
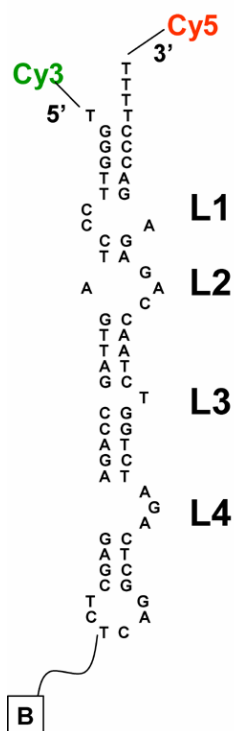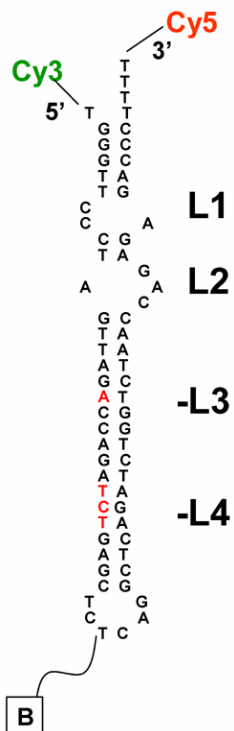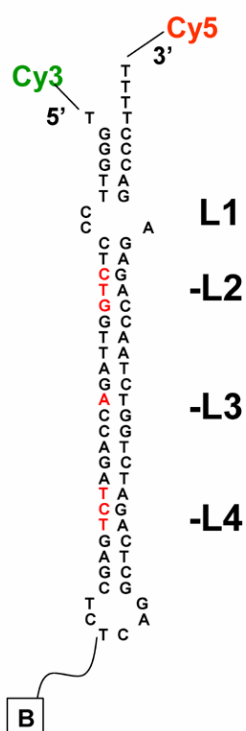
Scheme 2.2 The primary and secondary structures of TAR DNA and TAR DNA mutants used in the current experiments are illustrated. The DNAs were immobilized on the surface via a biotin linker (noted above as 'B') attached to a dT in the hairpin loop region.

22

Experimental Section

The HIV-1 NC protein used for these experiments was prepared as previously described.[20,36] Purified Cy3 (donor) and Cy5 (acceptor) labeled biotinylated DNA sequences were acquired from TriLink Biotechnologies (San Diego, CA). The MFOLD[37,38] predicted secondary structures are illustrated in Scheme 2.2. Dye-labeled DNA was immobilized using either biotinylated bovine serum albumen (BSA), or biotinylated polyethylene glycol (PEG).[20,35] The BSA method involves assembling predrilled polycarbonate films with an adhesive gasket (Grace Bio-Labs, Bend, OR) on top of clean coverslips. Inlet and outlet ports (Nanoport™, Upchurch Scientific, Oak Harbor, WA) were glued on top of the chambers. The chambers were incubated sequentially for 10 min with solutions of biotinylated BSA (Pierce Biotechnology, Inc., Rockport, IL; 2 mg/mL in distilled deionized water) and streptavidin (Molecular Probes, Eugene, OR); 0.2 mg/mL in HEPES buffer (*25 mM HEPES, pH 7.3*). The chambers were rinsed with distilled deionized water (50 μL) following each incubation step.

For the PEG immobilization, clean coverslips were treated with Vectabond/acetone 1% w/v solutions (Vector Laboratories, Burlingame, CA) for 5 min. Coverslips were then rinsed with H_2O and dried under a N_2 stream. The Vectabond coated coverslips were covered with templated silicone gaskets. The exposed areas were incubated with 25% m-PEG-SPA-2000 and 0.25% biotin-PEG-NHS-5000 (Nektar Therapeutics, San Carlos, CA) in 0.1 M sodium bicarbonate (Hyclone) for 3h. The silicone gaskets were removed, the excess PEG rinsed away, and the cover slips dried with N_2. The chamber was then treated with streptavidin and doubly-labeled oligomer

solution in the same manner as in the BSA method described above. Reactant solutions, which were delivered by syringe pumps,[20,35] contained HEPES buffer (*25 mM HEPES, pH 7.3*), various concentrations of $MgCl_2$ and NaCl, and an oxygen scavenger system (2-mercaptoethanol 1% v/v (Sigma-Aldrich, St. Louis, MO), β-D(+)glucose 3% w/v (Sigma-Aldrich, St. Louis, MO), glucose oxidase 0.1 mg/mL (Roche Applied Science, Hague Road, IN) and catalase 0.02 mg/mL (Roche Applied Science)).[39,40]

The scanning confocal optical/data collection system used the 514 nm line from an argon ion laser (model Reliant 150m, Laser Physics, Inc., West Jordan, UT).[20,35] The high numerical aperture oil immersion microscope objective (Zeiss Fluar, 100X, NA 1.3) was used to focus both the excitation and emission. Donor and acceptor fluorescence was collected with two avalanche photodiodes (APD) (Perkin Elmer Optoelectronics SPCM-AQR-15, Vaudreuil, Quebec, Canada). The intensity time trajectories were acquired with 1 ms time resolution. The collected donor and acceptor signals were corrected for background emission/noise and donor/acceptor crosstalk due to overlapping emission as previously described.[20] Also, ~1 ms blinking events (acceptor reversible photobleaching) were filtered from the data. Analysis of unfiltered data verified that the filtering process did not significantly distort the FRET dynamics on the time scales discussed in this paper. The corrected donor and acceptor intensities, $I_D(t)$ and $I_A(t)$, respectively, can be used to calculate directly the time trajectory of the apparent FRET efficiency, $E_A(t)$ as in:

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \qquad (1)$$

E_A(t) is related to the actual FRET efficiency, E_{FRET}(t), by the inclusion of the dye quantum efficiencies, ϕ_i, and detector quantum efficiencies, η_i according to:

$$E_{FRET}(t) = \frac{I_A(t)}{I_A(t) + I_D(t)\dfrac{\phi_A \eta_A}{\phi_D \eta_D}} \qquad (2)$$

In the case of the current experimental setup, it was determined that E_A(t) ~ E_{FRET}(t). E_A(t) autocorrelation curves, E_A Autocorrelation(τ), were obtained using the single molecule cross correlation curves, as detailed in a previous study,[20]

$$E_A Autocorrelation(\tau) = C(\tau)\left(-\frac{\langle I_D \rangle}{\langle I_A \rangle}\right) \qquad (3)$$

$$C(\tau) = \frac{\langle \delta I_D(t)\delta I_A(t+\tau)\rangle}{\langle I_D \rangle\langle I_A \rangle} = \frac{\langle I_D(t) I_A(t+\tau)\rangle}{\langle I_D \rangle\langle I_A \rangle} - 1 \qquad (4)$$

The data were fit in most cases to a single exponential decay, with autocorrelation lag time, amplitude, relaxation rate, and relaxation lifetime represented by τ, A_{AC}, λ, and τ', respectively.

For the simulations reported here, the donor and acceptor signals, I_D(t) and I_A(t), were simulated by using the kinetic Monte Carlo (KMC) method (For applications of KMC to FRET systems see, e.g.,[41]). If, for example, the system switches between states 1 and 2 with the FRET efficiencies E_1 and E_2, respectively, then the simulation proceeds as follows. First KMC is used to generate a stochastic "trajectory" switching between 1 and 2 according to the prescribed rate constants for the 1→2 and 2→1 transitions. While the molecule remains in one of the states, say state 1, the donor and acceptor signals are Poisson processes, with the probability of photon arrival per unit time equal to $I_0(1-E_1)$ and I_0E_1, respectively. Here I_0 is the observed intensity (the number of photons per unit time) from the donor in the absence of FRET. The above formula assumes the same

photodetection efficiency for both channels, which is the case for the current system. Again, a KMC procedure was used to generate random lists of photon arrival times obeying the above Poison processes. To account for the crosstalk between the donor and the acceptor channels, whenever the algorithm generated a donor photon arrival event, it was recorded as an acceptor photon with a specified probability (of 20%). Finally, background signal was generated for each channel as a Poisson process with the experimentally determined intensity.

Results and Discussion

TAR DNA hairpins in the absence of NC. To study the effects of NC on the secondary structure and dynamics of the donor/acceptor labeled TAR DNA mutants, we begin with a simple, well-defined case: SMFRET of the -L3-L4 TAR DNA mutant in the absence of NC. Previously, we have shown that this mutant is well described by a fully closed hairpin structure with a mean FRET value of $<E_A> = 0.96$ and with undetectable E_A fluctuations.[24] A typical $E_A(t)$ trajectory for –L3-L4 mutants is shown in Figure 2.1A. The durations of these trajectories are limited by photobleaching of Cy5.

One of the simplest ways to represent SMFRET data is in an ensemble single molecule E_A histogram, where ensemble refers to the combination of single molecule E_A histograms from many hairpins under the same conditions. In order to construct ensemble E_A histograms, the individual $E_A(t)$ trajectories were boxcar time-averaged, or "smoothed" with alternative bin times, τ_B, of 3 ms, 25 ms, and 250 ms. In boxcar averaging a group of N adjacent $E_A(t)$ points are averaged together where $N=\tau_B /\tau_D$. Here τ_D is the time spacing in the original non-time averaged data (i.e. 1 ms). As τ_B is increased the signal-to-noise ratio of the data increases due to time averaging. FRET fluctuations are, however, also smoothed, or averaged out of the data by this process, if they occur on a faster time scale than τ_B. The boxcar time-averaged single molecule trajectories were combined to yield ensemble E_A histograms as shown in Figure 2.1B. The $\tau_B =3$ ms E_A histogram for this hairpin shows a single peak due to the **C** form (see Scheme 2.1). The width of the peak is primarily due to photon shot noise rather than E_A fluctuations, which is verified by E_A simulations (see below).
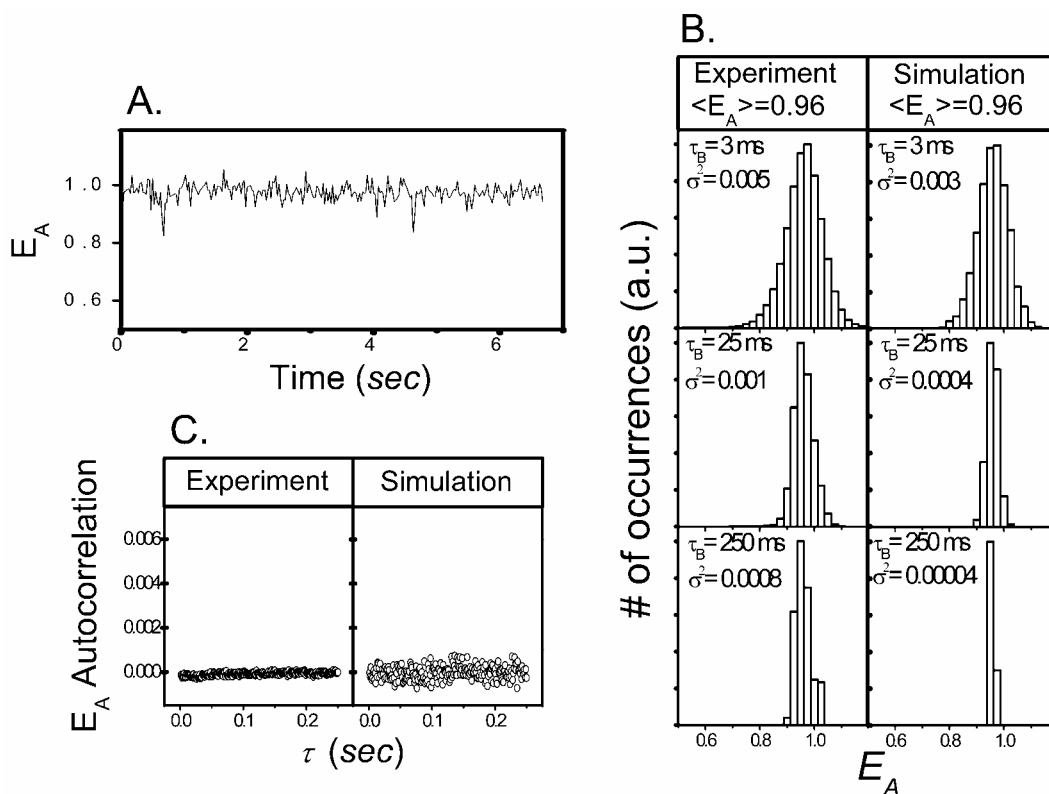
Figure 2.1 SMFRET of –L3-L4 TAR DNA. (A) A representative single molecule E_A trajectory, with a bin time, τ_B, of 25 ms. (B) Experimental E_A values presented in the form of ensemble E_A histograms with the data binned into different τ_B (left panels). Monte Carlo simulated ensemble E_A histograms, based on a single-state system, are included for comparison (right panels). (C) The 1-ms experimental ensemble E_A trajectory is presented in the form of an autocorrelation (left panel). The simulated curve is shown for comparison (right panel).

Simulated histograms, based on a single-state system with photon shot noise, are shown in Figure 2.1B (right panels) at each respective τ_B, for comparison. The small differences in variance between the single state model and experimental values:

$$\Delta\sigma^2 = \left|\sigma^2_{exp} - \sigma^2_{theory}\right| \tag{5}$$

reflect experimental error rather than actual E_A fluctuations. A careful comparison of individual trajectories suggests that small errors in establishing the background level in the donor and acceptor channels are the source of this error. Thus, for the 250 ms τ_B histogram in Figure 2.1B the small residual error we observed, σ^2, will be denoted by σ^2_{back} to reflect the source of the error.

Ensemble E_A autocorrelation is another way to characterize E_A dynamics, where ensemble refers to the averaging of single molecule E_A autocorrelation curves from many hairpins under the same conditions. For TAR DNA hairpins in the presence of NC, the E_A autocorrelation exhibits exponential decays (see below) due to secondary structure fluctuations. For -L3-L4 in the absence of NC, no detectible component is observed within the experimental detection limit, i.e., $A_{AC} < 0.0005$ (Figure 2.1C, left panels). A system that exhibits A_{AC} below this detection limit can be considered a relatively "static" single state secondary structure. A simulated single molecule autocorrelation is also shown in Figure 2.1C (right panel). The undetectably small A_{AC} indicates that the -L3-L4 TAR DNA molecules in the absence of NC do not fluctuate in their structure. This static **C** form was observed for each TAR DNA mutant studied in the absence of NC at each concentration of Mg^{2+}.

SMFRET of the -L2-L3-L4 TAR DNA/NC Complex. Figure 2.2 shows NC-induced structural fluctuations for the -L2-L3-L4 TAR DNA mutant that only retains a single loop near the top of the stem (see Scheme 2.2). The shift in $<E_A>$ from the NC-free value of 0.96 (data not shown) to a value of 0.87 in the presence of 445 nM NC and 0.2 mM Mg^{2+} (Figure 2.2, column 1) is due to the presence of the open form at equilibrium.[20] Previous studies have suggested that a two-state equilibrium model can describe the fluctuations in short DNA hairpins.[42,43] In the current work, the experimental E_A histogram and autocorrelation data (Figure 2.2, column 1) are in near quantitative agreement with a two-state kinetic simulation (Figure 2.2, column 2) using a kinetic Monte Carlo (KMC) algorithm that is described in the Experimental Section and elsewhere.[41] The small differences in variance σ^2, between the simulation and experimental values in Figure 2.2 reflect experimental error σ^2_{back} rather than actual E_A fluctuations. As in the case of the NC free results (see Figure 2.1) the variance σ^2 for all of the experimental τ_B=250 ms E_A histograms in Figure 2.2 is not significantly larger than that expected from background error σ^2_{back}.

This implies that the opening and closing dynamics are both homogenous and simple first-order within experimental error. Thus, the data do not indicate the presence of either long-lived intermediates or multiple opening/closing pathways for the -L2-L3-L4 TAR DNA/NC Complex.

The required parameters for the simulation (Table 2.1) were determined by the following procedure. First, the experimental normalized E_A autocorrelation decay was fit by analytical results for a two-state interconversion as follows:

$$\frac{<E_A(t)E_A(t+\tau)>}{<E_A(t)>^2} - 1 = \frac{(E_{A,open}-E_{A,close})^2\, k_{open}k_{close}}{(k_{open}E_{A,open}+k_{close}E_{A,close})^2}\, e^{-\lambda\tau} = \frac{(E_{A,open}-E_{A,close})^2\, K_{eq}}{(K_{eq}E_{A,open}+E_{A,close})^2}\, e^{-\lambda\tau} \quad (6)$$

Where λ, the rate constant, is simply $\lambda = k_{open}+k_{close}$. E_{close} was determined to be 0.96 from the experiment in the absence of NC. In the presence of NC, $<E_A>$ satisfies Equation 7 for a two-state system:

$$<E_A> = \frac{K_{eq}\cdot E_{A,open}+1\cdot E_{A,close}}{1+K_{eq}} \quad (7)$$

Equations 6 and 7 were used to calculate K_{eq} and E_{open} based on the experimentally obtained A_{AC} and $<E_A>$. The rate constants, k_{open} and k_{close}, can thus be determined from fitting the E_A autocorrelation with a single decay constant, λ, and K_{eq} value.

Experimental data were also acquired for 1.0 mM (Figure 2.2, column 3) and 2.5 mM Mg^{2+} (Figure 2.2, column 4), showing a shift in the equilibrium (see Table 2.1) toward the closed secondary structure as the Mg^{2+} concentration is increased.
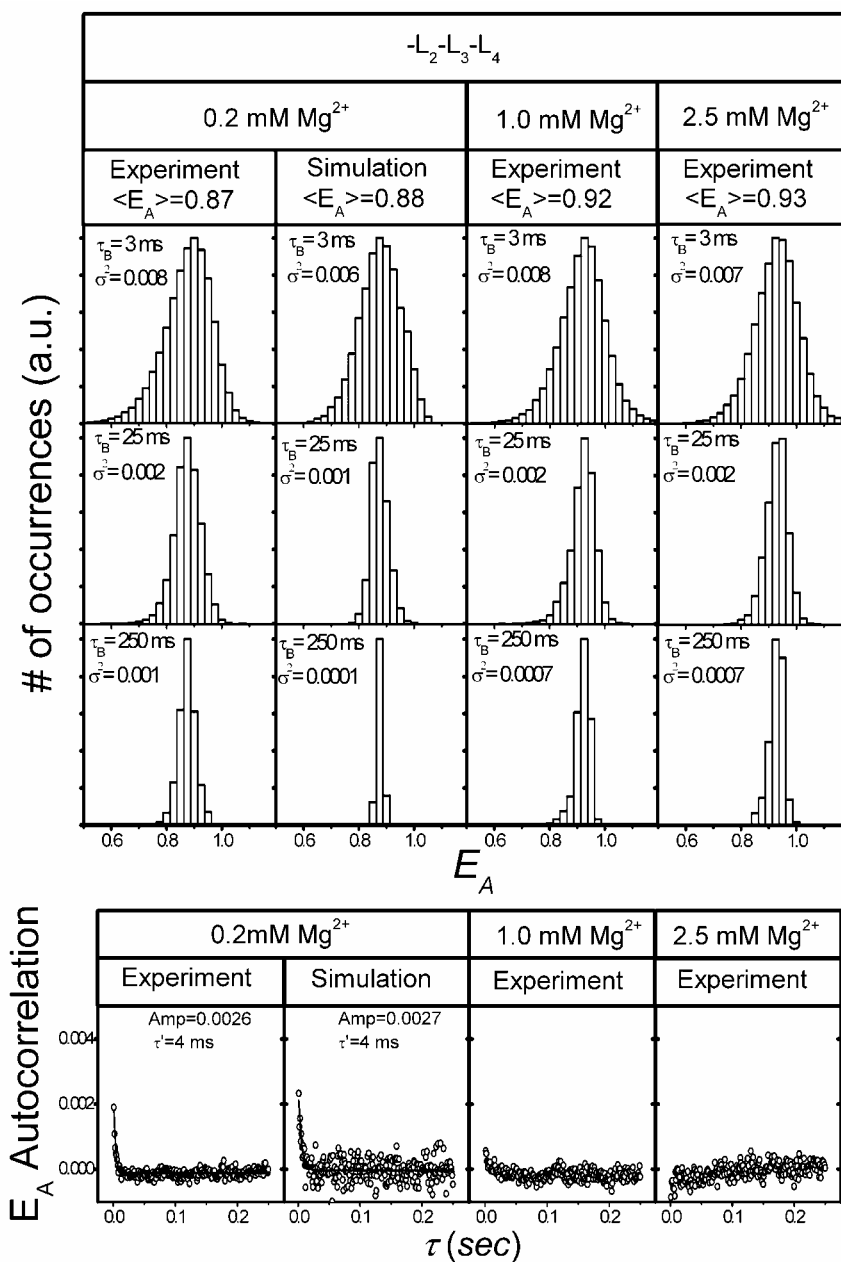
Figure 2.2 (top) Experimentally obtained ensemble E_A histograms for the -L2-L3-L4 TAR DNA mutant in the presence of 445 nM NC and three different Mg^{2+} concentrations. A simulated distribution is shown for the 0.2 mM Mg^{2+} data. (bottom) E_A autocorrelation curves are shown for the same conditions.

Table 2.1 Experimental SMFRET Parameters for the TAR DNA Mutants in the Presence of NC, and Associated Parameters Determined by Fitting the SMFRET Data Using the Two-State Interconversion Model .

Sequences	-L3-L4	-L2-L3-L4/NC			-L3-L4/NC		
	0.2 mM Mg^{2+}	0.2 mM Mg^{2+}	1.0 mM Mg^{2+}	2.5 mM Mg^{2+}	0.2 mM Mg^{2+}	2.5 mM Mg^{2+}	5.0 mM Mg^{2+}
$E_{A,close}$	0.96	0.96	0.96	0.96	0.96	0.96	0.96
$<E_A>$	0.96	0.88	0.92	0.93	0.77	0.89	0.92
A_{AC}	-	0.0026	-	-	0.0070	0.0032	0.0012
Rate (ms^{-1})	-	0.24	-	-	0.18	0.084	0.13
K$_{eq}$	-	5.23	0.69	0.46	11.9	2.34	1.63
$E_{A,open}$	-	0.86	0.86*	0.86*	0.75	0.85	0.90
k$_{open}$(ms^{-1})	-	0.20	-	-	0.17	0.059	0.078
k$_{close}$(ms^{-1})	-	0.039	-	-	0.014	0.025	0.048

* The A_{AC} for these measurements was too small to fit, thus the -L2-L3-L4/NC/0.2 mM Mg^{2+} values were used to determine K$_{eq}$. The inability to measure A_{AC} for these cases may be due to opening/closing rates that are too rapid to resolve by the experimental methods.

SMFRET of -L3-L4 TAR DNA/NC complex. E_A data for NC induced dynamics of the -L3-L4 TAR DNA hairpin are shown in Figure 2.3. By analogy to the -L2-L3-L4 TAR DNA data, higher concentrations of Mg^{2+} shift the equilibrium distribution toward the **C** form. However, $<E_A>$ for this mutant is significantly smaller than that observed for the -L2-L3-L4 mutant under similar conditions. This is consistent with opening through the L1-L2 regions. The experimental E_A histograms for different $[Mg^{2+}]$ and τ_B are considerably broader than the best-fit two-state simulation of the data. The excess broadening in the experimental E_A data is even present at relatively long times as evidenced by the large σ^2 in the $\tau_B =250$ ms histogram.

The presence in Figure 2.3 of broadening that significantly exceeds experimental error for the TAR DNA -L3-L4 mutant in the 250 ms τ_B histogram indicates that the opening/closing process has a significant probability of occurring at times longer than 250 ms. This implies that the opening/closing is not a simple first order process because the dominant relaxation time for the E_A fluctuations is on the 3-10 ms time scale. Qualitatively, the excess broadening indicates non-single-exponential end-to-end distance dynamics, and, correspondingly, an underlying complexity in the secondary structure dynamics. For example, one or more long-lived intermediates along the reaction coordinate for the -L3-L4 mutant may exist, leading to two or more additional rate constants. Alternately, the complex secondary structure dynamics may be a reflection of parallel pathways for opening and closing kinetics, i.e., heterogeneous opening/closing kinetics. Hypothetical assignments for the non-exponential dynamics will be presented
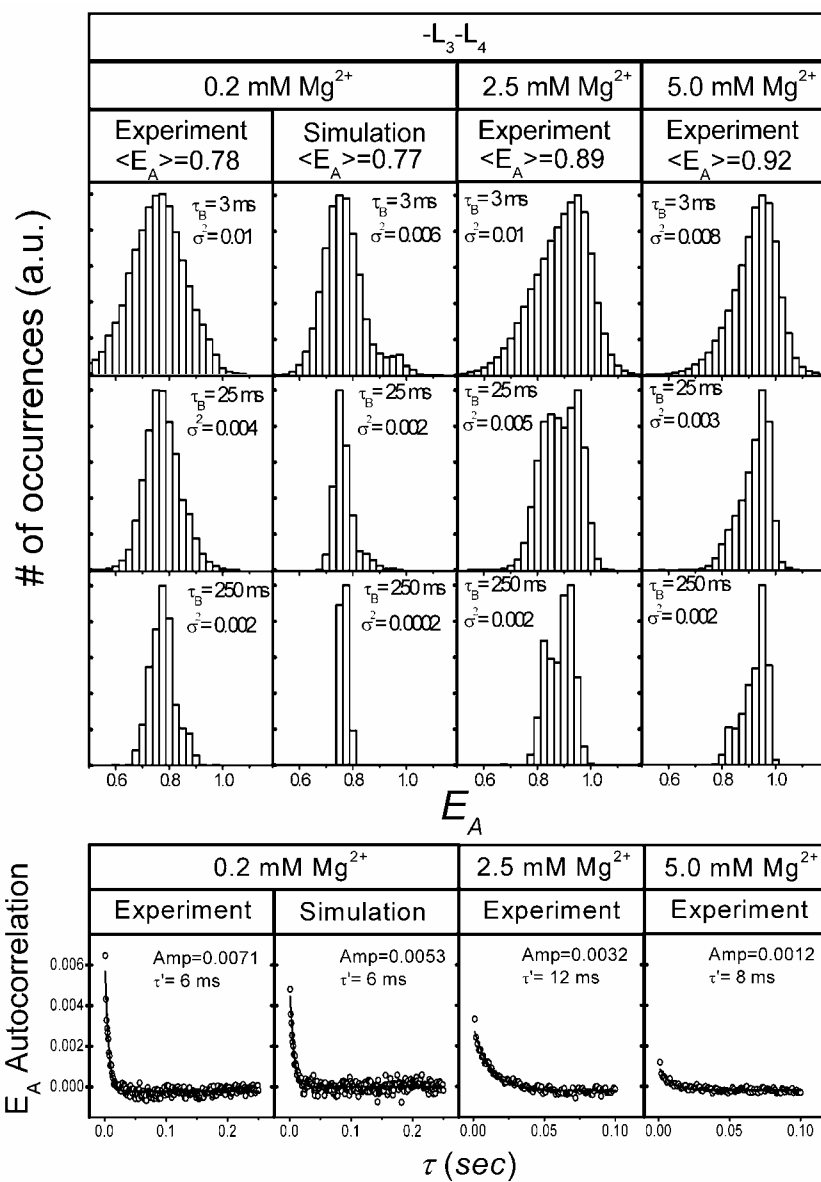
Figure 2.3 (top) Ensemble E_A histograms for the -L3-L4 TAR DNA mutant in the presence of 445 nM NC are presented for three different Mg^{2+} concentrations. The simulated values at 0.2 mM Mg^{2+} are included for comparison. (bottom) E_A autocorrelation decays are compared for the three reaction conditions.

below. Furthermore, heterogeneity due to surface immobilization or chemical damage is assumed to be negligible in this system, consistent with the observations that analogous heterogeneity is observed regardless of the immobilization strategy and for various mutants and NC preparations. This is analogous to the heterogeneity that has been observed for several other nucleic acid/protein systems.[25-27]

The trend in A_{AC} vs. K_{eq} also demonstrates clear deviation from a simple two-state equilibrium model (Table 2.1). As shown previously for other TAR DNA constructs[20], the effect of increasing $[Mg^{2+}]$ is to shift the equilibrium towards an equal distribution of **C** and **Y** forms, as demonstrated by the dependence of K_{eq} on $[Mg^{2+}]$ shown in Table 2.1. A simple two-state equilibrium model predicts that A_{AC} reaches a maximum value at an equal distribution of **C** and **Y** forms, which clearly deviates from the experimental results. In fact, the data shown in Table 2.1 and Figure 2.3 reveal that A_{AC} decreased when the equilibrium was shifted towards an equal distribution of **C** and **Y** forms. The experimental deviation from a simple two-state equilibrium model prediction for the dependence of K_{eq} on $[Mg^{2+}]$ trends is suggestive of more than one intermediate along the reaction coordinate. However, other explanations might be able to account for the observed deviation from a simple two state model, such as a salt effect on the chain stiffness, which could alter the end-to-end distance and FRET values. Another indication that a simple two state model does not adequately describe the opening/closing process of -L3-L4 TAR DNA in the presence of NC is the significant increase in the best-fit $E_{A,open}$ values in Table 2.1 as a function of increasing Mg^{2+}. The apparent decrease in $E_{A,open}$ as a function of decreasing Mg^{2+} apparently indicates that more than one open secondary

structure of the hairpin is present at equilibrium. This furthermore implies that the experimental values obtained from the two state analyses for -L3-L4 TAR DNA are qualitative at best since a 3 state or even more complex treatment would be necessary to accurately describe this system.

Further studies of the NC's effects on -L3-L4 TAR DNA structural dynamics were performed as a function of NC concentration (Figure 2.4). In all cases, the -L3-L4 mutant exhibits heterogeneity on both short (5 ms) and long (>250 ms) time scales.

As shown in Figures 2.3 and 2.4, the heterogeneous opening/closing dynamics in the -L3-L4 mutant were observed over a large range of NC and Mg^{2+} concentrations. The dependence of $<E_A>$ on the concentration of NC and Mg^{2+} are compared in Figure 2.5. The overall effects of NC and Mg^{2+} on the relative population of the **Y** form are in opposition to each other. Further experiments as a function of $[Na^+]$ (data not shown) exhibited similar effects as that of Mg^{2+}, although higher $[Na^+]$ were necessary to achieve comparable E_A values.
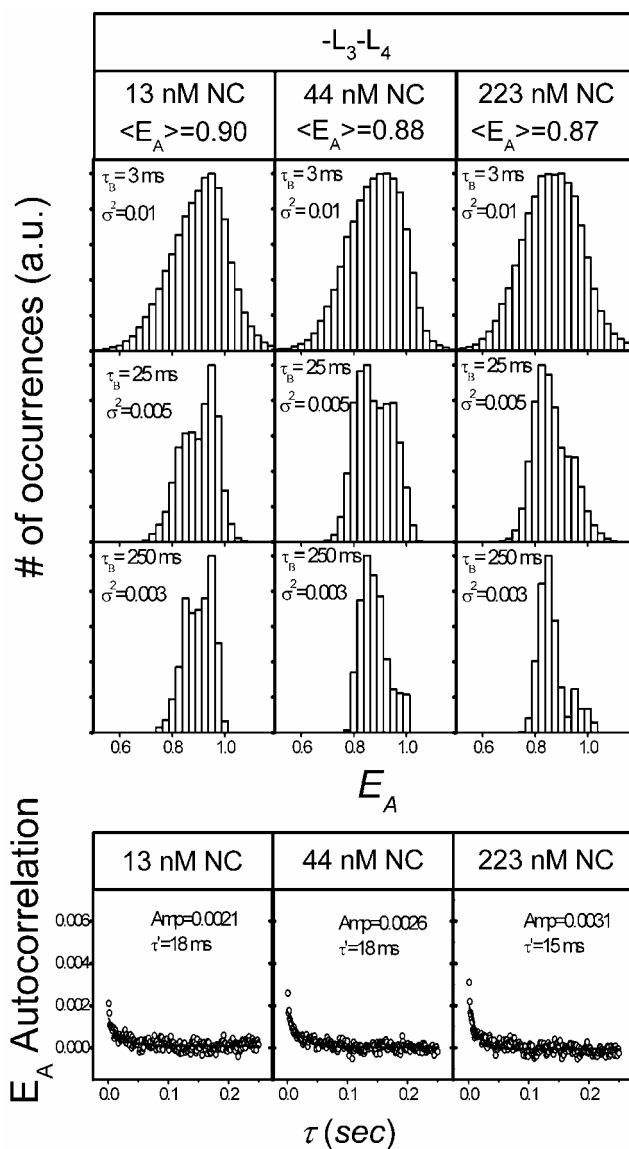
Figure 2.4 (top) Ensemble E_A histograms for the -L3-L4 mutant at various NC concentrations are compared. (bottom) The E_A autocorrelations for each ensemble are shown. All experiments were performed in the presence of 0.2 mM Mg^{2+} and 40 mM NaCl.
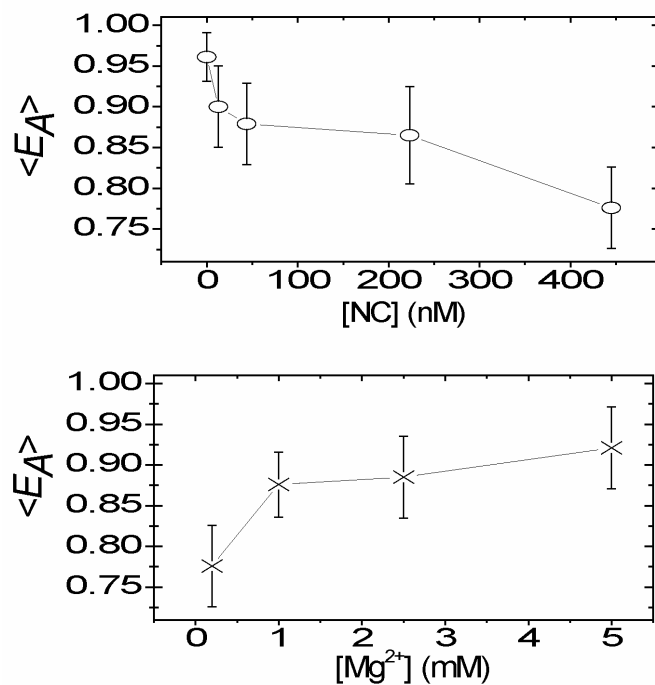
Figure 2.5 The trends in $\langle E_A \rangle$ as a function of [NC] (top) and [Mg^{2+}] (bottom) are shown. The [NC] dependent experiments were performed at 0.2 mM Mg^{2+}, and the [Mg^{2+}] dependent experiments were performed at 445 nM NC. The lines are included as guides. The error bars denote σ at τ_B =250 ms for each experiment.

Further evidence that the opening/closing dynamics in the -L3-L4 mutant are heterogeneous is presented in Figure 2.6. Here the ensemble E_A histograms for the -L3-L4 mutant hairpins in the presence of 44 nM NC and 0.2 mM Mg^{2+} have been sorted and summed with respect to the single molecule E_A values. The resultant "sub-ensemble" E_A histograms are plotted in color in the top-panel in Figure 2.6. The corresponding sub-ensemble autocorrelation curves are shown at the bottom-of Figure 2.6. The black "sub-ensemble" is a fully closed population with $<E_A> = 0.98$, while the red and blue sub-ensembles are partially open structures with $<E_A>$ values of 0.87 and 0.83, respectively.

The sub-ensemble of hairpins with the highest average E_A values (coded black in Figure 2.6) has FRET values that correspond to a fully closed structure, apparently the **C** form. The autocorrelation curve for this sub-ensemble shows no measurable structural dynamics (i.e., $A_{AC} < 5 \times 10^{-4}$) indicating a relatively static hairpin structure with very little DNA end-to-end distance fluctuations. It is unlikely that this sub-ensemble is due to hairpins with no bound NC, since at an NC concentration of 44 nM, approximately 2.5 NC molecules are estimated to be bound per hairpin assuming $K_d \sim 100$ nM and 8 independent binding sites per hairpin. Furthermore, NC/hairpin collisions should occur on a shorter timescale than the >250 ms lifetime of the sub-ensemble of hairpins with the highest average E_A values, ensuring multiple NC binding/disassociation events per second.

In contrast, the E_A autocorrelation curves for the partially open "red" and "blue" sub-ensembles show clear evidence of large amplitude end-to-end distance fluctuation on a range of time scales from 2-17 ms. The "blue" sub-ensemble exhibits the most rapid

40

dynamics, on the ~ 2 ms time scale near our instrumental resolution. The E_A autocorrelation times for the "red" sub-ensemble are much longer and the decay is clearly non-single exponential.

It is important to emphasize that the basic assumption of this analysis is that the different sub-ensembles are a consequence of kinetic heterogeneity on a > 250 ms time scale due to long-lived secondary structures and/or NC binding states. As described above, analogous heterogeneity has been observed for a broad range of NC and Mg^{2+} concentrations, *e.g.*, the 250 ms histograms in Figure 2.4.

Mechanistic implications of the SMFRET data. The observed SMFRET data suggest that the NC-induced destabilization in the L1-L2 regions of TAR DNA is in equilibrium between a relatively static structure with very little end-to-end distance fluctuations and a much more dynamic structure. Further, the structural fluctuations occur on multiple time scales, indicating a complex, multidimensional relationship between bound NC and the secondary structure in the L1-L2 regions of TAR DNA. Despite the extensive new observations, the available experimental data of various types on this process is still not sufficient to unequivocally determine the minimum number of kinetic species that are required to adequately describe this process. Perhaps, even more importantly, it is not possible for a given kinetic intermediate to specify the secondary structures, numbers of bound NC molecules, or the location of the bound NC molecules. Nevertheless, it is useful to consider simple kinetic models for the opening/closing process in an attempt to begin to identify possible physical origins of the observed effects.
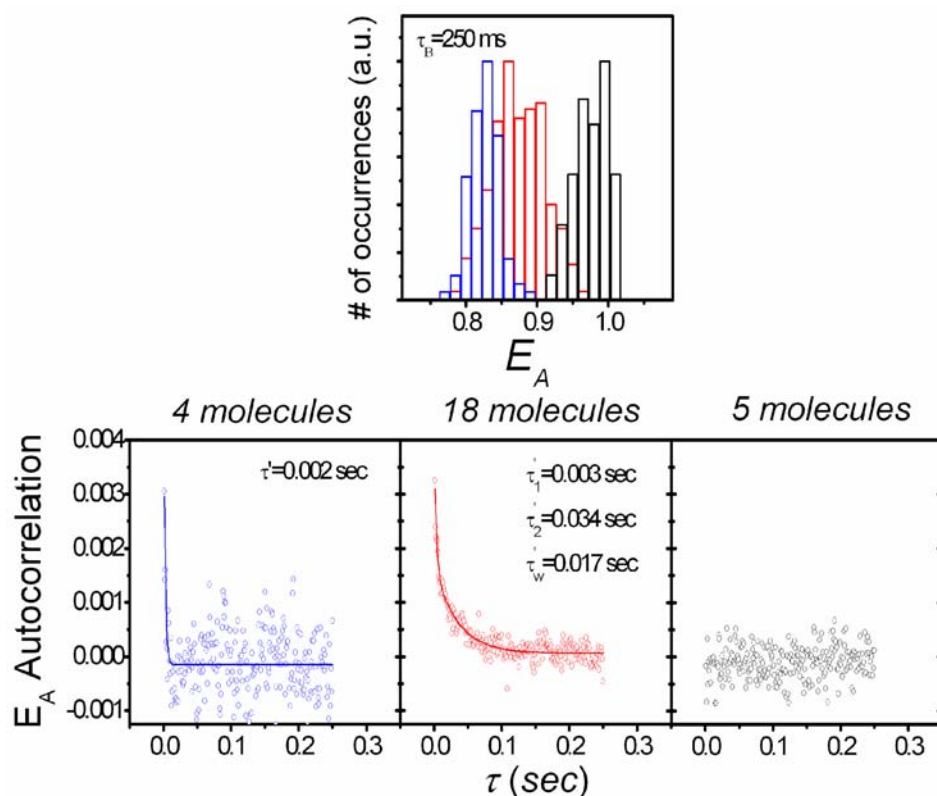
Figure 2.6 (Top) The 250 ms τ_B ensemble E_A histogram is shown for -L3-L4 TAR DNA, with 44 nM NC and 0.2 mM Mg^{2+}. The molecules have been sorted into primarily closed (black), primarily open (blue) and intermediate (red) structures based on their single molecule E_A values. (Bottom) The corresponding sub-ensemble E_A autocorrelation curves are shown, with the respective exponential curves and fitting parameters for each set. The A_{AC} curves for the intermediate molecules required double-exponential fitting. Each individual component, and the weighted time constant, τ'_w, are indicated in the plot.

The data in Figure 2.6 demonstrates that there is a significant population of long-lived closed hairpins, suggesting a kinetic model in which the initial opening event is rate-limiting, as follows:

$$C \xrightleftharpoons[k_{-1}]{k_1} Y_1 \xrightleftharpoons[k_{-2}]{k_2} Y_2 \qquad (8)$$

where k_1, the rate constant for initiating hairpin opening, may be as slow as 1/250 ms based on the SMFRET data. Within this model the millisecond time-scale dynamics observed in Figures 2.4 and 2.6 are assigned to secondary structure dynamics among the partially open **Y** forms. The relatively slow initiation of NC induced TAR DNA opening suggests that the transition state for opening might involve thermally activated breaking of the Watson-Crick base pairs in the L1-L2 region, with only a partial lowering of the barrier for opening by the bound NC.

Interestingly, the model implied by Equation 8 is in contrast with the two-state **C/Y** interconversion analysis of the -L2-L3-L4 TAR DNA data described above, since the latter analysis suggests a simple **C/Y** interconversion on the millisecond timescale. It is conceivable, however, that the model in Equation 8 could also apply to the -L2-L3-L4 TAR DNA hairpin if both of the two states were assigned to different **Y** forms. This would require that very little **C** form be present at equilibrium for this hairpin in the presence of NC.

An alternative, equally plausible class of models that could account for the heterogeneous SMFRET dynamics is described by the following scheme:

$$C' \xrightarrow[k_{-1}]{k_1} C'' \xrightarrow[k_{-2}]{k_2} Y' \xrightarrow[k_{-3}]{k_3} Y'' \qquad (9)$$
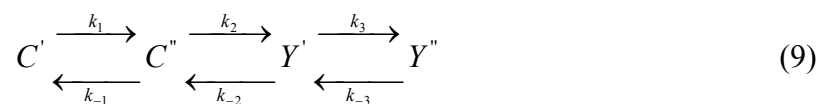
in which there are two **C** forms that differ by secondary structure and/or NC binding number and configuration, leading to very different open rates. This model allows for both a long-lived **C'** form and short lived **C''** form. For example **C''** might be "activated" by an NC bound in a location, *e.g.*, a bulge, that promotes the opening process. Thus, this model preserves the basic assignment of the SMFRET dynamics to the opening/closing process but still allows a sub-population of static **C** forms.

At this point the data do not offer a means for distinguishing between the alternate kinetic models. Clearly more research will be necessary to determine whether either of these models captures the salient features of the NC induced opening/closing dynamics of TAR DNA hairpins, or whether even more elaborate kinetic treatments will be necessary.

Conclusions

SMFRET was used to study NC-induced fluctuations of various DNA hairpins over a broader range of conditions and in more depth than in previous studies. As observed in other single molecule studies of nucleic acid structural fluctuations such as the hairpin ribozyme,[44] secondary structure dynamics over a broad distribution of time scales are observed. τ_B-dependent histograms of E_A and autocorrelation analysis reveal complex dynamics that occur on two distinct time scales when both L1-L2 are present in the hairpin. In contrast, when L2 is absent, the fluctuations are constrained to a ~5 ms lifetime. The experiments herein reveal a complex mechanism for secondary structure fluctuations, with dynamic processes occurring over a wide time range, i.e., ~5 to >250 milliseconds, and with the involvement of long-lived intermediates.

Acknowledgements

References

(1)   Draper DE, Reynaldo LP (1999) *Nucl. Acids Res.* 27: 381-388.

(2)   Stockley PG (1999) in *Frontiers Mol. Biol.*, ed. Baumberg S, Vol. 21, pp. 22-58.

(3)   Doyle M, Jantsch MF (2002) *J. Struct. Biol.* 140: 147-153.

(4)   Salas M, Freire R, Soengas MS, Esteban JA, Mendez J, Bravo A, Serrano M, Blasco MA, Lazaro JM, Blanco L, Gutierrez C, Hermoso JM (1995) *FEMS Microbio. Rev.* 17: 73-82.

(5)   Sevenich FW, Langowski J, Weiss V, Rippe K (1998) *Nucl. Acids Res.* 26: 1373-1381.

(6)   Schultz A, Langowski J, Rippe K (2000) *J. Mol. Biol.* 300: 709-725.

(7)   Telesnitsky A, Goff SP (1997) in *Retroviruses*, ed. Varmus HE (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 121-160.

(8)   Levin JG, Guo J, Rouzina I, Musier-Forsyth K (2005) in *Prog. Nucleic Acid Res. Mol. Biol.*, Vol. 80, pp. 217-286.

(9)   Zhang Y, Barklis E (1995) *J. Virol.* 69: 5716-22.

(10)   Zhang Y, Qian H, Love Z, Barklis E (1998) *J. Virol.* 72: 1782.

(11)   Feng YX, Campbell S, Harvin D, Ehresmann B, Ehresmann C, Rein A (1999) *J. Virol.* 73: 4251-6.

(12)   Chan B, Weidemaier K, Yip W-T, Barbara PF, Musier-Forsyth K (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96: 459-464.

(13)   Wu T, Guo J, Bess J, Henderson LE, Levin JG (1999) *J. Virol.* 73: 4794-805.

(14)   Guo J, Wu T, Anderson J, Kane BF, Johnson DG, Gorelick RJ, Henderson LE, Levin JG (2000) *J. Virol.* 74: 8980-8.

(15)   You JC, McHenry CS (1994) *J. Biol. Chem.* 269: 31491.

(16)   Rein A, Henderson LE, Levin JG (1998) *Trends Biochem. Sci.* 23: 297-301.

(17)   Johnson PE, Turner RB, Wu ZR, Hairston L, Guo J, Levin JG, Summers MF (2000) *Biochemistry* 39: 9084-91.

(18)   Azoulay J, Clamme JP, Darlix JL, Roques BP, Mély Y (2003) *J. Mol. Biol.* 326: 691-700.

(19)   Beltz H, Azoulay J, Bernacchi S, Clamme JP, Ficheux D, Roques B, Darlix JL, Mély Y (2003) *J. Mol. Biol.* 328: 95-108.

(20)   Cosa G, Harbron EJ, Zeng Y, Liu HW, O'Connor DB, Eta-Hosokawa C, Musier-Forsyth K, Barbara PF (2004) *Biophys. J.* 87: 2759-67.

(21)   Urbaneja MA, Wu M, Casas-Finet JR, Karpel RL (2002) *J. Mol. Biol.* 318: 749-64.

(22)   Williams MC, Rouzina I, Bloomfield VA (2002) *Acc. Chem. Res.* 35: 159-66.

(23)   Lu HP, Iakoucheva LM, Ackerman EJ (2001) *J. Am. Chem. Soc.* 123: 9184-9185.

(24)   Zhuang X, Bartley LE, Babcock HP, Russell R, Ha T, Herschlag D, Chu S (2000) *Science* 288: 2048.

(25)   Rothwell PJ, Berger S, Kensch O, Felekyan S, Antonik M, Wohrl BM, Restle T, Goody RS, Seidel CAM (2003) *Proc. Natl. Acad. Sci. U.S.A.* 100: 1655.

(26)   Ha T (2004) *Biochemistry* 43: 4055-4063.

(27)   Ha T, Rasnik I, Cheng W, Babcock HP, Gauss G, Lohman TM, Chu S (2002) *Nature* 419: 638-641.

(28)   Xie Z, Srividya N, Sosnick TR, Pan T, Scherer NF (2004) *Proc. Natl. Acad. Sci. U.S.A.* 101: 534-539.

(29)   Tan E, Wilson TJ, Nahas MK, Clegg RM, Lilley DMJ, Ha T (2003) *Proc. Natl. Acad. Sci. U.S.A.* 100: 9302.

(30)   Lu PH, Xun L, Xie XS (1998) *Science* 282: 1877.

(31)   Zhuang X, Rief M (2003) *Curr. Opin. Struct. Biol.* 13: 88-97.

(32)   Onoa B, Tinoco I, Jr. (2004) *Curr. Opin. Struct. Biol.* 14: 374-379.

(33)   Onoa B, Dumont S, Liphardt J, Smith SB, Tinoco I, Jr., Bustamante C (2003) *Science* 299: 1892-1896.

(34)   Previous work, using fluorescence correlation spectroscopy to study a similar cTAR DNA and truncated NC system, reported conformational fluctuations on the microsecond time scale. (See refs. 18-19) This technique, because it is limited by the diffusion-controlled sampling lifetime, is not capable of measuring slow dynamics, such as those reported here, that occur on millisecond (and much longer) time scales.

(35)   Liu HW, Cosa G, Landes CF, Zeng Y, Mullen DG, Barany G, Musier-Forsyth K, Barbara PF (2005) *Biophys. J.* submitted.

(36)   Lee BM, De Guzman RN, Turner BG, Tjandra N, Summers MF (1998) *J Mol Biol* 279: 633-49.

(37)   Zuker M (2003) *Nucl. Acids. Res.* 31: 3406-3415.

(38)   Zuker M, Mathews DH, Turner DH (1999) *Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide* (Kluwer Academic Publishers, Dordrecht).

(39)   Ha T (2001) *Curr. Opin. Struct. Biol.* 11: 287-292.

(40)   Harada Y, Sakurada K, Aoki T, Thomas DD, Yanagida T (1990) *J. Mol. Biol.* 216: 49-68.

(41)   Wang Z, Makarov DE (2003) *J. Phys. Chem. B* 107: 5617-5622.

(42)   Bonnet G, Krichevsky O, Libchaber A (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95: 8602-8606.

(43)   Ritort F, Bustamante C, Tinoco JR I (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99: 13544-13548.

(44)   Rueda D, Bokinsky G, Rhodes MM, Rust MJ, Zhuang X, Walter NG (2004) *Proc. Natl. Acad. Sci. U.S.A.* 101: 10066-10071.

Chapter 3: Probing Nucleation, Reverse Annealing, and Chaperone Function Along the Reaction Path of HIV-1 Single Strand Transfer

Introduction

Reverse transcription of the HIV-1 RNA genome involves several critical nucleic acid rearrangement steps that are chaperoned (catalyzed) by the HIV-1 nucleocapsid protein, NC.[1,2] Mechanistic investigations of these rearrangements have been challenging due to the extreme structural and kinetic heterogeneity of these processes, which are believed to involve a heterogeneous distribution of nucleic acid/protein complexes of variable composition and unknown secondary structure.[3] These challenges have been particularly well-documented for the NC *annealing* of the transactivation response region (TAR) RNA of the HIV-1 genome to the complementary sequence (TAR DNA) in minus-strand strong-stop DNA.[4] Liu *et al*. recently investigated the kinetics of an *in vitro* model (eqn (i) in Figure 3.1) for this multi-step reaction, i.e. the annealing of two hairpins to produce the thermodynamically favored duplex.[5] In the Liu *et al*. experiments isolated TAR hairpins were immobilized on a biologically compatibilized cover-slip that was located within a flow system that supplied "fresh" unaggregated solution of complementary DNA or RNA hairpin. Real-time fluorescence single molecule spectroscopy (SMS) of oligonucleotides labeled with donor (D) and acceptor (A) dyes was used to simultaneously monitor the annealing kinetics, the secondary structure of the reactants, and the state-of-aggregation of the hairpins, *in situ*. Internal distances between

51

labeled sites in reacting hairpin pairs were probed by time-resolved fluorescence resonance energy transfer, i.e. SM-FRET, between the donor and acceptor dyes.

Based on the data of Liu *et al.* and previous observations, these authors proposed a hypothetical mechanism for the annealing reaction that is summarized by eqn (ii) and (iii) in Figure 3.1. The reacting hairpins in this mechanism undergo a rapid NC-induced equilibrium between closed and partially melted "Y" shaped conformations, which is supported by previous experiments.[6,7] The chaperone (catalytic) activity of NC is envisioned as arising from essentially two independent effects: a NC-induced partial melting of the Watson-Crick pairing of the reactant hairpins,[8-11] and a NC-induced decrease in the energy cost of bringing the hairpins together to form the encounter complex, due presumably to a screening of the negative charges on the hairpins and perhaps through specific interactions,[12-16] The nucleation of annealing occurs in an *encounter complex* that is comprised of partially melted TAR DNA and TAR RNA hairpins and several bound molecules of NC.[5] Two *nucleation-complex* isomers are envisioned for the annealing process, depending on whether the annealing occurs in the region of the L1L2 internal loops or, alternatively, in the region of the L4 internal bulge and hairpin loop (HL) of TAR DNA. The L1-L4 and HL regions are indicated in eqn (i) in Figure 3.1.
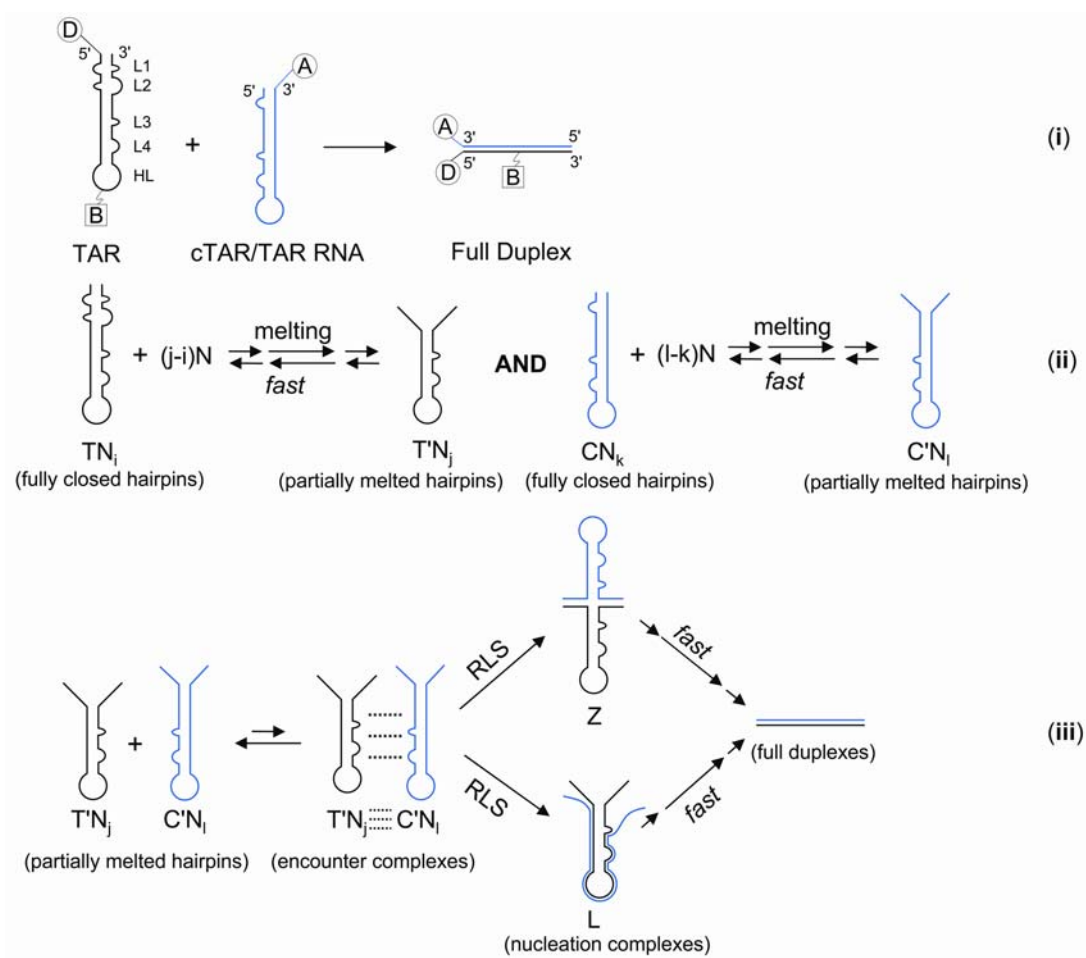
Figure 3.1   A hypothetic kinetic scheme for NC chaperoned annealing of dye-labeled Cy3-TAR DNA to its Cy5-labeled complements, with Cy3 as fluorescence donor (D) and Cy5 acceptor(A). Here, T denotes TAR DNA, and C denotes complementary cTAR DNA or TAR RNA. The term N denotes the nucleocapsid protein, NC. In this scheme, N binds to T and C, leading to a partially melted structure, namely the Y form of T (T') and C(C'). The subscripts, i, j, k and l are used to describe the number of NC bound to nucleotides. Two partially melted hairpins form an encounter complex that leads to the formation of nucleation complexes. The annealing can go through either zipper nucleation or loop nucleation, therefore, forming zipper nucleation complexes (Z) or loop nucleation complexes (L) both leading to the formation of fully annealed duplexes.
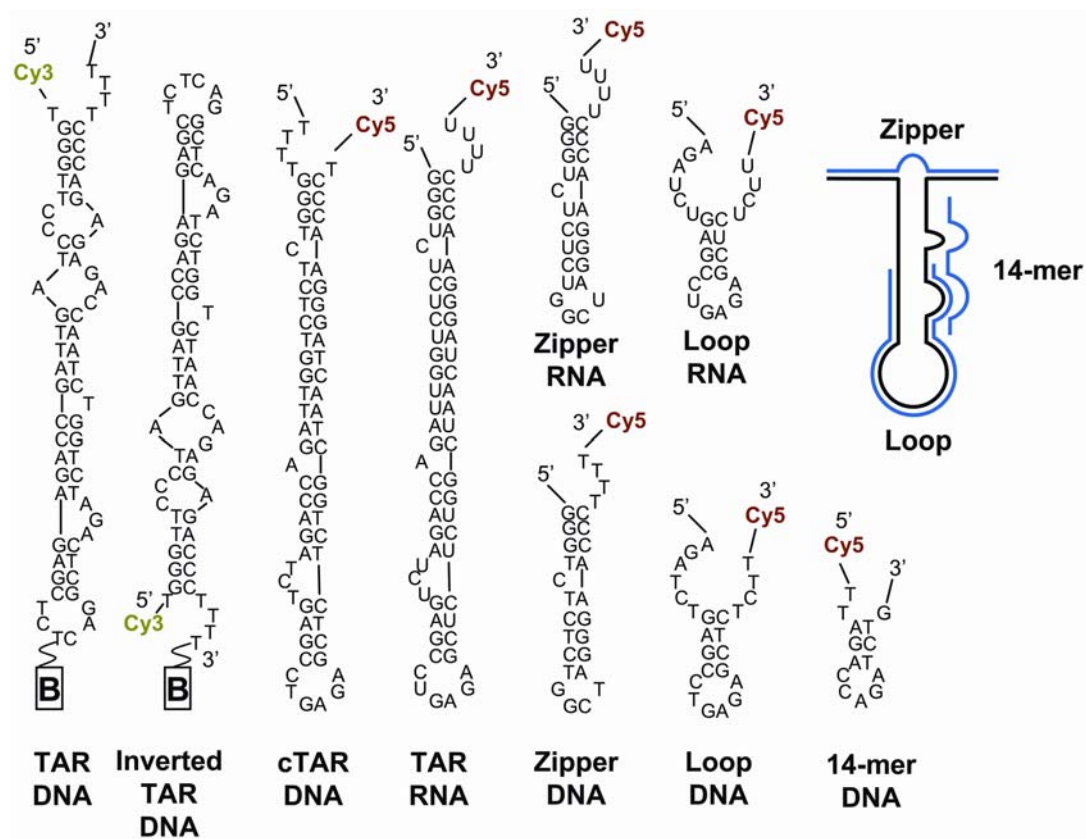
Figure 3.2 The structures of various oligonucleotides used. The secondary structures are predicted by mfold program (http://www.bioinfo.rpi.edu/applications/mfold/old/dna/) [17].

54

Herein the annealing reaction between TAR DNA and various DNA or RNA oligonucleotides (Figure 3.2) is probed by a new SMS approach in which single molecule kinetic data are acquired while the immobilized TAR DNA hairpins are exposed to a time programmed concentration sequence of different targeted oligonucleotides with and without the NC chaperone present in the solution. This procedure chemically "drags" individual pairs of reacting hairpins through the reactant states, intermediate states and back again.[5] This approach offers new information on the different stages of the annealing mechanism, especially the putative nucleation complexes. The experiments in this paper have been built on previous simpler SMS TAR DNA annealing studies with DNA oligonucleotides.[5]

Experimental Section

SMS data in this paper were recorded by repetitive confocal scanning imaging as described previously.[5,18] HIV-1 NC was synthesized as previously described.[5] Various functionalized DNA hairpins (all purchased from TriLink Biotechnologies, San Diego, CA) and RNA hairpins (purchased from Dharmacon RNA Technologies, Lafayette, CO) were used without further purification as previously described.[5] TAR DNA was immobilized on the coverslip of an assembled flow chamber. Three syringe pumps delivered three solutions: NC, target complementary hairpins (containing Mg^{2+}) and buffer solution (containing buffer A and Mg^{2+}) separately. All the solutions contained buffer A (40mM NaCl, 25 mM HEPES, pH 7.3 and glucose oxygen scavenger system).[5]

Results and Discussion

The multi-component, oligonucleotide SMS experiments that are the focus of this paper can be put in context by first considering the previously investigated annealing of "full length" hairpin cTAR DNA to an immobilized TAR DNA hairpin (eqn (i), Figure 3.1).[5,18] Single molecule kinetic results for this reaction are shown in the first column of Figure 3.3. Here the irreversible NC catalyzed annealing reaction was initiated by exposing a dilute, immobilized sample of the Cy3-TAR DNA to a freshly mixed solution of Cy5-cTAR DNA and NC at zero time, $t = 0$, in analogy to a stopped-flow experiment. The freshly mixed solution was prepared *in situ* in a mixing chamber which combined various concentrations of acceptor labeled short oligonucleotides solutions in HEPES buffer, and a 889 nM solution of NC. Prior to $t = 0$ only a buffer solution was flowed into the flow-cell. The acceptor and donor intensities, $I_A(t)$ and $I_D(t)$, were recorded by sample scanning confocal microscopy at various stages of the reaction correspondingly, and the donor and acceptor images analyzed. For each immobilized hairpin, an apparent FRET efficiency, E_A, was determined as follows:

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \tag{1}$$

By measuring E_A for each hairpin, at various times, t, after introducing the complementary Cy5-hairpin solution FRET trajectories (Figure 3.3A) were recorded to monitor the instantaneous distance between the 5' end of the immobilized Cy3-TAR DNA hairpin and the 3' end of the Cy5-hairpins.
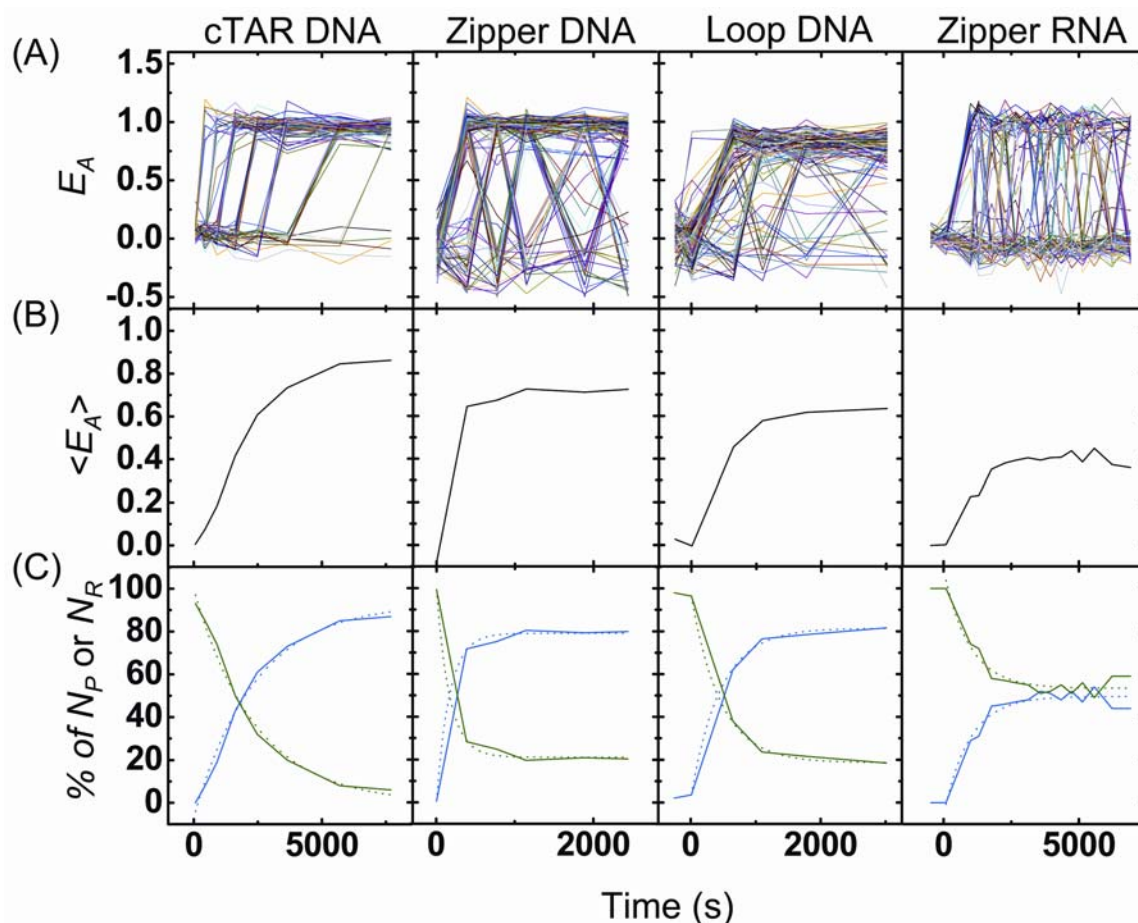
Figure 3.3 Room temperature single-molecule kinetic measurement on the annealing of Cy3-TAR DNA to 2 nM Cy5-cTAR, 25 nM Cy5-zipper DNA, 25 nM loop DNA and 50 nM Cy5-zipper RNA at 0.2 mM Mg^{2+} and 889 nM NC. Top panels show the E_A of each single molecule during the reaction, where each colored line corresponds to a single molecule. Middle panels show the ensemble mean E_A increases during the annealing reaction. Bottom panels show the number of product (N_P, shown as % among total number of molecules) molecules during the annealing reaction (blue, solid), while the number of reactant (green solid) (N_R) molecules decrease by the same amount. Both of the increase/decrease trends over time can be fitted with a single exponential (dotted lines). The data was collected from 100 single molecules for cTAR annealing kinetic measurement, 140 for zipper annealing and 200 for loop annealing.

FRET trajectories for individual reacting pairs (Figure 3.3A) show the previously reported behavior of exhibiting discrete switching at the time of annealing from a low FRET value due to an isolated Cy3-TAR DNA hairpin, to a high value due to an annealed Cy3-TAR DNA/Cy5-cTAR pair. FRET is efficient for the annealed pair due to the close proximity of the 5' end of TAR DNA and the 3' end of cTAR. Previously reported histogram of the FRET from many hairpins at different times during annealing show the expected dual peaks, corresponding to the two-state annealing reaction.[18] Very few of the hairpins reveal *reverse annealing* (high to low FRET changes) indicating that the annealing reaction highly favors the full duplex product state. This also demonstrates that photobleaching,[19-21] which is also manifested by a similar high to low FRET change, is not a significant complication in these experiments. Some hairpins do not anneal during the time period due to non-idealities such as unlabeled cTAR and clustering or imperfect immobilization. Due to these imperfections the mean FRET value as a function of time (Figure 3.3B) does not perfectly reflect the true FRET changes, since it reaches an asymptotic value of 0.86, while the E_A of the annealed form is ~ 1. Figure 3.3C portrays the number of surviving reactant hairpins N_R (green curve) and the number of annealed product N_P (blue curve) pairs, respectively. These were determined by counting the number of hairpins that were below/above an E_A threshold of 0.4, as described elsewhere.[18]

The second and third column in Figure 3.3 correspond to annealing experiments with shorter DNA oligonucleotides that are targeted for "zipper" (L1L2 stem loops) and "loop" (L3, HL) regions of TAR DNA, as shown in Figure 3.2. SMS data on annealing

59

of both targeted oligonucleotides show evidence of *reversible annealing* that leads at later times to an equilibrium distribution of annealed and unannealed TAR. For example the individual E_A trajectories show much more high-to-low FRET transitions than the TAR DNA/cTAR case. The $<E_A>$ curves and the number of reactant and product data are also consistent with an equilibrium mixture at long times, with an apparent dissociation constant K_d of ~ 10 nM and ~ 16 nM for the zipper and loop DNA, respectively. One can estimate K_d from the ~ 70% and ~ 60% annealing percentage at 25 nM concentration for zipper and loop DNA, respectively. The annealed adducts of TAR DNA with the target oligonucleotides are arguably *models for intermediates in the annealing reaction* of full length cTAR or TAR RNA and indeed, in minus strand transfer itself. We envision that annealed form of TAR DNA with zipper is a model for the putative "Z" nucleation-complex in Figure 3.1, and, in turn, the annealed adduct of TAR DNA with the loop oligonucleotide, is a model for the "L" nucleation-complex in Figure 3.1. To simplify the nomenclature in this paper, the equilibrium form of the annealed adduct of two nucleic acids will be denoted as follows, nucleic-acid-1/nucleic-acid-2, e.g. TAR DNA/zipper DNA. The annealing of TAR DNA with TAR RNA, zipper RNA and loop RNA are more difficult to study accurately due to their higher tendency of forming aggregates and sticking to surface, but generally exhibit similar kinetic behavior to the DNA oligonucleotides. As an example, the fourth column of Figure 3.3 shows the kinetics of TAR DNA/zipper RNA annealing.
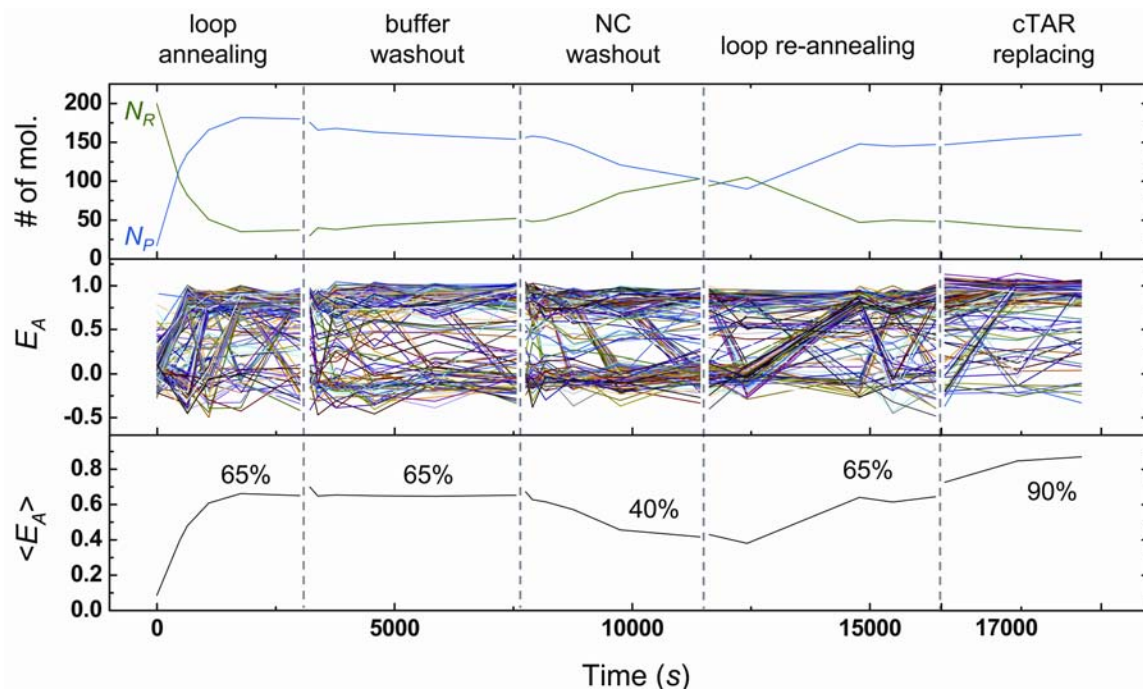
Figure 3.4 An example of room temperature single-molecule kinetic measurement of the annealing reaction between immobilized Cy3-TAR DNA and 25 nM Cy5-loop DNA in the presence of NC (889 nM). The sample was then washed out by buffer and NC sequentially, re-annealed with 25 nM Cy5-labeled loop DNA (re-annealing) and reacted with 25 nM Cy5-cTAR DNA (cTAR). All solutions contained 0.2 mM Mg^{2+}. Top panels show the number of annealed (blue) and un-annealed TAR DNA (green) molecules at different reaction stages. Middle panels show the single molecule E_A trajectories, where each colored line corresponds to a single molecule. Bottom panels show the mean E_A of the single molecules in the middle panel. The final annealed percentages in each reaction stage are also shown in the corresponding graphs.

In order to characterize the nucleic acid rearrangement pathways available to the TAR DNA/targeted-oligonucleotide adducts we subjected these models for nucleation-complexes to a sequence of solutions containing buffer only, buffer plus NC, and finally, a solution of NC plus cTAR. Typical results are shown in Figure 3.4 for the loop DNA oligonucleotide. In the first time epoch of the experiment, immobilized TAR was exposed to a loop DNA plus NC solution, leading to an annealing equilibrium. This was followed by a period in which the loop DNA plus buffer solution was rapidly replaced with a buffer only solution. The major effect of the buffer-only period was to freeze the concentration of the TAR/loop DNA adduct, even though the equilibrium constant strongly favors dislocated adducts in the absence of oligonucleotides in solution. In contrast, when NC was added to the solution (in the third epoch) the concentration of the adduct decreased relatively rapidly and continuously.

These data, therefore, demonstrate that NC not only catalyzes the forward annealing process, but also catalyzes the reverse annealing process. This is expected according to microscopic reversibility for a NC chaperoned process. Since NC interacts more strongly with single stranded DNA than with double stranded DNA, the combined NC effects strongly suggest that the transition state for the annealing reaction possesses more "single stranded character" than the reactants, products, and even the stable intermediates along the reaction path. This is qualitatively consistent with the proposed "Z" nucleation-complex structure in Figure 3.1. The final epoch in Figure 3.4 involves a NC-promoted strand displacement of loop DNA by cTAR. This step in the sequence is a simple assay on the "activity" of the immobilized TAR hairpins, allowing for a validation

62

Table 3.1 Rate constants for annealing (k_a) and reverse annealing (k_r) at 0.2 mM Mg^{2+} in the presence or absence of 889 nM NC.

Parameter[a]		cTAR DNA	Zipper DNA	Loop DNA[c]	TAR RNA	Zipper RNA	Loop RNA[c]
k_a $(10^5 s^{-1} M^{-1})$	- NC	0	0	0	0	0	0
	+ NC	8[b]	10	1	2	0.1	0.06
k_r $(10^{-4} s^{-1})$	- NC	0	1[d]	0.1[d]	0	5	0
	+ NC	0	18	4	0	5	3

Notes:

[a] k_a denotes the annealing rate constant and k_r denotes the reverse annealing rate constant. The annealing reaction kinetic curve is measured in the presence (+NC) or absence (-NC) of NC and fitted by a single exponential function. The single exponential rate constant k is $k = k_r + k_a \cdot [NA]$ (NA = reacting nucleic acid). k_r is determined by washing out the annealed product with NC (+NC) or buffer only (-NC) solution, and fitting the kinetic curve by a single exponential function with a rate constant k_r. k_a is determined from k and k_r as $k_a = (k - k_r)/[NA]$.
[b] value from our previous published data.[18]
[c] The loop annealing reaction kinetics is potentially complicated by steric interference of the biotin/streptavidin attachment group (denoted by "B" in Figure 3.1). To avoid this complication, loop annealing kinetic were measured with an inverted TAR DNA construct, with the B group attached to the T64. Although there does not appear to be appreciable steric interference for the TAR DNA-loop DNA annealing,[5] there is measurable steric interference of the TAR DNA-loop RNA annealing.
[d] This value is based on the slow decay component of Np. The fast decay component is associated with reverse annealing of 1-armed annealed TAR DNA/zipper or loop DNA adduct.

of the entire procedure. Since efficient cTAR annealing was observed, it is unlikely that the TAR DNA hairpins were damaged or poorly immobilized by the programmed sequence of reagents. Very similar results were observed for the annealing of TAR DNA to zipper DNA, zipper RNA and loop RNA.

Various observables and derived kinetic parameters for the annealing reaction of TAR DNA with the various nucleotides are listed in Table 3.1. Some clear trends are apparent in the data. For example NC is clearly required for both the annealing and reverse annealing reactions to be rapid. Also, the reverse annealing of the full length oligonucleotides (cTAR and TAR RNA) are much slower than the short oligonucleotides consistent with the idea that the annealed adduct of the TAR DNA with the short oligonucleotides is indeed a model for the nucleation complex for the full length annealing reactions. In other words, considerably more base pairs must be broken to achieve the proposed transition state for reverse annealing of the full length oligonucleotides. Presumably, the reverse annealing of the TAR DNA/zipper DNA adduct is analogous to the slow reverse reaction process in Figure 3.1 that converts "Z" to "TN==CN", demonstrating reversibility in this step of the mechanism for the first time.

Further evidence for the reversibility of this pathway was obtained by forming an equilibrium mixture of the TAR DNA/zipper DNA adduct by NC annealing and then rapidly replacing the NC plus zipper DNA solution with a NC plus cTAR (dye labeled) solution. The zipper DNA hairpins were observed to be efficiently replaced by the cTAR (data not shown). Presumably, this net "strand displacement" reaction occurred due to NC induced annealing between cTAR and the unannealed TAR DNA. (The unannealed

TAR DNA presumably resulted from NC induced reverse annealing of the TAR DNA/zipper-DNA adduct.) Various strand displacement reactions were undertaken as summarized in Table 3.2. In every case where a stable annealed product between TAR DNA and a full length oligonucleotide was expected, the strand displacement reaction was too slow to observe. For example, it was observed that the cTAR in a TAR DNA/cTAR adduct were efficiently displaced by a concentrated mixed solution of NC and zipper DNA (and also not washed-out by NC only). Correspondingly, for all adducts of TAR DNA with short oligonucleotides, efficient strand displacements of the short oligonucleotide by another oligonucleotide were observed.

Further insights into the annealing mechanism were obtained by a different type of application of strand displacement, involving a short Cy5-labeled DNA oligonucleotide that is complementary to L3L4 (see 14-mer in Figure 3.2). We used this 14-mer DNA as a "dynamic probe" for base paring in the TAR DNA L3L4 region during the irreversible annealing of cTAR (Figure 3.5). Due to the small number of base pairs in the TAR/14-mer adduct its reverse annealing rate is very rapid. This was demonstrated by the single molecule FRET trajectories in the top panel of Figure 3.5. Starting with a TAR DNA sample with no complementary oligonucleotide (epoch I), a solution of NC and Cy5-labeled 14-mer was added at the beginning of epoch II. Complete annealing was observed within the mixing time of the flow system, followed by rapid reversible on-off events in E_A trajectories due to rapid annealing/reverse annealing. The rapid NC induced annealing is also reflected in the rapid rise of the mean FRET as a function of

Table 3.2 Summary of strand-displacement experiments.[a]

REACTION SEQUENCE
------ **Efficient Strand Displacement** ------
Cy3-TAR/Cy5-Zipper-DNA+ Cy5-cTAR
Cy3-TAR/Cy5-Loop-DNA + Cy5-cTAR
Cy3-TAR/Cy5-Zipper-RNA + Cy5-cTAR
Cy3-TAR/Cy5-Loop-RNA + Cy5-cTAR
Cy3-TAR/Cy5-14-mer-DNA + cTAR
------ **Inefficient Strand Displacement** ------
Cy3-TAR/cTAR + Cy5-cTAR
Cy3-TAR/cTAR + Cy5-Loop-DNA
Cy3-TAR/cTAR + Cy5-cTAR + Cy5-14-mer-DNA

[a] The annealing and replacement experiments were run in buffer containing 0.2 mM Mg^{2+} and 889 nM NC, 25 nM Cy5-zipper/loop nucleotide and Cy5-cTAR DNA with the exception that the Cy5-14-mer DNA measurement which was run with 50 nM Cy5-14-mer DNA and 2.5 nM unlabeled cTAR.
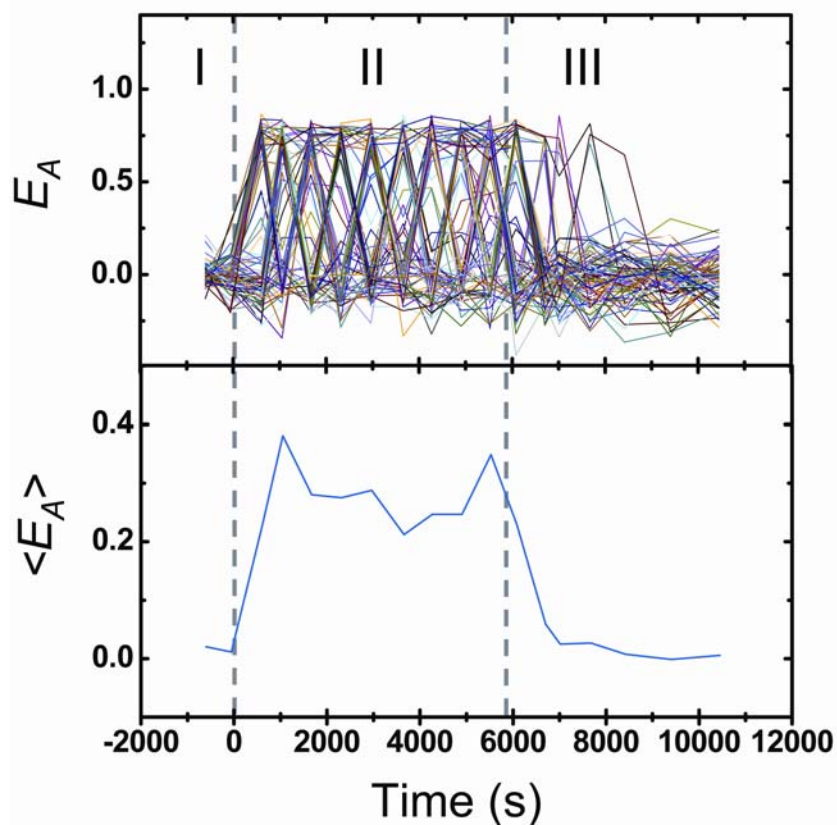
Figure 3.5 Room temperature single molecule E_A trajectories (top) and ensemble mean E_A (bottom) during the annealing reaction of Cy5-14-mer DNA annealed to Cy3-TAR DNA then replaced by non-labeled cTAR DNA. Three time epochs during the reaction are plotted: (I) no annealing occurred in the presence of 50 nM Cy5-14-mer DNA only, (II) reversible annealing reaction occurred in the presence of 50 nM Cy5-14-mer DNA and 889 nM NC, and (III) the annealed Cy5-14-mer DNA was displaced by non-labeled cTAR DNA in the presence of 2.5 nM non-labeled cTAR DNA, 50 nM Cy5-14-mer DNA, and 889 nM NC. All reactions were run in the presence of 0.2 mM Mg^{2+}.

time curve (bottom pane, epoch II). The most interesting part of Figure 3.5 is the epoch III. At the beginning of epoch III, cTAR DNA was added to the NC and Cy5-labeled 14-mer solution, and a rapid irreversible displacement of the 14-mer by cTAR was observed at the usual time scale for TAR DNA/cTAR annealing. The fact that the irreversible annealing of non-labeled cTAR DNA with TAR occurs with a concomitant displacement of the 14-mer is highly consistent with the hypothetical mechanism in Figure 3.1, which assumes full annealing occurs soon after the rate limiting nucleation step.

While the main features of the SMS data are well explained by the mechanism in Figure 3.1, certain small but discernable features in the data, suggest that the actual mechanisms may be somewhat more complex. For example, consider the interruption of the annealing equilibrium of the TAR DNA/zipper DNA adduct with a buffer-only washout step, as shown in Figure 3.6. As expected from the mechanism in Figure 2.1, the population of the adduct decayed only very slowly after the buffer entered the cell, i.e. the reverse annealing step requires NC to occur rapidly. The E_A trajectories also exhibit the expected underlying behavior, i.e. very few transitions between the low and high FRET levels after the buffer solution enters the sample cell. But, unexpectedly, immediately after the TAR/zipper DNA adduct was exposed to a buffer-only solution, a small but statistically significant drop was observed in the number of TAR DNA/zipper DNA adducts (N_P). A corresponding, unexpected small decrease in $<E_A>$ was also observed. Both effects were reproduced in several trials. We tentatively assign the small, rapid drop in N_P to rapid reverse annealing of a conformational sub-state of the TAR/zipper adduct that is only partially annealed, and as a result can be washed out by

the incoming buffer solution. A good prospect for this sub-state is a "one-arm" annealed form of the TAR DNA/zipper DNA adduct, as outlined in the equation on the top of Figure 3.6. The "one-arm" annealed form is analogous to the previously reported rapidly reversible annealing of TAR DNA with a 13-mer short DNA oligonucleotide,[5] which is complementary to only one arm of the "Y" form of TAR. This non-ideality does not appear to be a factor for the TAR DNA/loop DNA adduct or for the full length TAR DNA/cTAR reaction, itself. Of course, all the above observations are subject to our time resolution (> 120 sec). The faster events beyond the time resolution will be missed and the actual mechanism could be more complex.
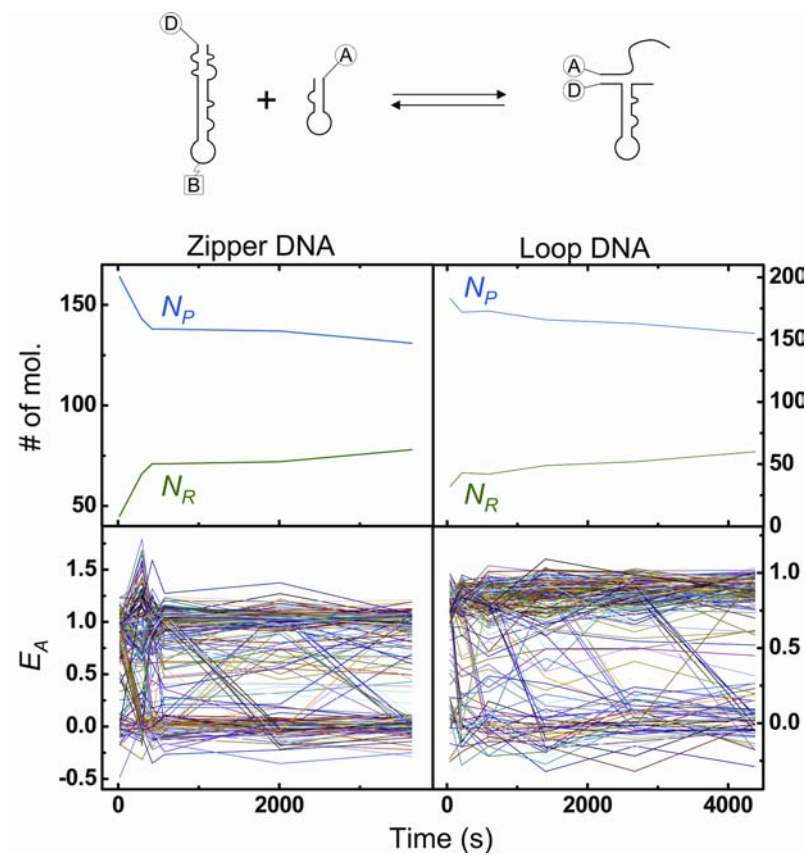
Figure 3.6 Comparison between single molecule measurements on buffer washout (0.2 mM Mg^{2+}) after immobilized Cy3-TAR DNA + Cy5-zipper (left column) and immobilized inverted Cy3-TAR DNA + Cy5-loop DNA annealing reaction (right column). Top panels show the number of annealed (blue) and un-annealed molecules during washout and bottom panels show the single-molecule E_A trajectories. The equation above panels shows a hypothetic scheme for the formation of a one-arm annealed TAR DNA/zipper DNA adduct.

Conclusions

A new SMS approach has been developed for analyzing the mechanism of protein induced nucleic acid rearrangements that involve multiple oligonucleotides and are highly heterogeneous and complex. The approach involves exposing one of the oligonucleotides (which is immobilized) to a sequence of solutions, containing the complementary oligonucleotide, other target oligonucleotides, buffer-only, and the chaperone protein in various combinations. This procedure effectively drags the nucleic acid protein systems from reactants through key intermediates, and finally toward the rearranged products, while monitoring the conformational states and dynamics of the system with SM-FRET. Herein this sequence has been applied to investigate the NC protein chaperoned annealing of TAR DNA to cTAR DNA, which is an *in vitro* model for the minus strand transfer step in HIV-1 reverse transcription. The new results strongly suggest that the nucleation event for annealing of transactivation response region TAR DNA to cTAR DNA involves base pair formation, in different regions of TAR DNA. In addition, the results clearly demonstrate that NC induces *reversible* annealing at various stages along the reaction path of the annealing reaction.

Acknowledgments

References

(1)   Levin JG, Guo J, Rouzina I, Musier-Forsyth K (2005) *Prog Nucleic Acid Res Mol Biol* 80: 217-286.

(2)   Darlix J-L, Lapadat-Tapolsky M, de Rocquigny H, Roques BP (1995) *J Mol Biol* 254: 523-37.

(3)   Rothwell PJ, Berger S, Kensch O, Felekyan S, Antonik M, Wohrl BM, Restle T, Goody RS, Seidel CAM (2003) *Proc Natl Acad Sci USA* 100: 1655-1660.

(4)   Hong MK, Harbron EJ, O'Connor DB, Guo J, Barbara PF, Levin JG, Musier-Forsyth K (2003) *J Mol Biol* 325: 1-10.

(5)   Liu H-W, Cosa G, Landes CF, Zeng Y, Mullen DG, Barany G, Musier-Forsyth K, Barbara PF (2005) *Biophys J* 89: 3470-3479.

(6)   Cosa G, Harbron EJ, Zeng Y, Liu H-W, O'Connor DB, Eta-Hosokawa C, Musier-Forsyth K, Barbara PF (2004) *Biophys J* 87: 2759-2767.

(7)   Cosa G, Zeng Y, Liu H-W, Landes CF, Makarov DE, Musier-Forsyth K, Barbara PF (2006) *J Phys Chem B* 110: 2419-2426.

(8)   Hargittai MRS, Gorelick RJ, Rouzina I, Musier-Forsyth K (2004) *J Mol Biol* 337: 951-968.

(9)   Urbaneja MA, Wu M, Casas-Finet JR, Karpel RL (2002) *J Mol Biol* 318: 749-764.

(10)   Williams MC, Rouzina I, Wenner JR, Gorelick RJ, Musier-Forsyth K, Bloomfield VA (2001) *Proc Natl Acad Sci USA* 98: 6121-6126.

(11)   Williams MC, Gorelick RJ, Musier-Forsyth K (2002) *Proc Natl Acad Sci USA* 99: 8614-8619.

(12)   Stoylov SP, Vuilleumier C, Stoylova E, De Rocquigny H, Roques BP, Gerard D, Mely Y (1997) *Biopolymers* 41: 301-312.

(13)   Le Cam E, Coulaud D, Delain E, Petitjean P, Roques BP, Gerard D, Stoylova E, Vuilleumier C, Stoylov SP, Mely Y (1998) *Biopolymers* 45: 217-229.

(14)   Lapadat-Tapolsky M, De Rocquigny H, Van Gent D, Roques B, Plasterk R, Darlix JL (1993) *Nucleic Acids Res* 21: 831-839.

(15)   Stoylov SP, Stoylova E, Todorov R, Schmiedel P, Thunig C, Hoffmann H, Roques BP, Le Cam E, Coulaud D, Delain E, Gerard D, Mely Y (1999) *Colloids and Surfaces, A: Physicochemical and Engineering Aspects* 152: 263-274.

(16)   Fisher RJ, Fivash MJ, Stephen AG, Hagan NA, Shenoy SR, Medaglia MV, Smith LR, Worthy KM, Simpson JT, Shoemaker R, McNitt KL, Johnson DG, Hixson CV, Gorelick RJ, Fabris D, Henderson LE, Rein A (2006) *Nucleic Acids Res* 34: 472-484.

(17)   Zuker M (2003) *Nucleic Acids Res* 31: 3406-3415.

(18)   Liu H-W, Zeng Y, Landes CF, Kim YJ, Zhu Y, Ma X, Vo M-N, Musier-Forsyth K, Barbara PF (2007) *Proc Natl Acad Sci USA* 14: 5261-5267.

(19)   Widengren J, Schwille P (2000) *J Phys Chem A* 104: 6416-6428.

(20)   Fuereder-Kitzmueller E, Hesse J, Ebner A, Gruber HJ, Schuetz GJ (2005) *Chem Phys Lett* 404: 13-18.

(21)   Sabanayagam CR, Eid JS, Meller A (2005) *J Chem Phys* 123: 224708/1-7.

Chapter 4: Single Molecule Study of the Inhibition of HIV-1 Transactivation Response Region DNA:DNA Annealing by Argininamide

Introduction

In the minus-strand transfer of HIV-1 reverse transcription, the transactivation response region (TAR) RNA of the genome is annealed to the complementary "TAR DNA" in minus-strand strong-stop DNA via a multi-step process that is chaperoned by the nucleocapsid protein (NC). A molecular-level understanding of the minus-strand transfer process of HIV-1 has emerged in recent years.[1-10] Several mechanistic insights into these processes have been obtained by applying single molecule spectroscopy (SMS) to an *in vitro* model for minus-strand transfer, which is portrayed in Figure 4.1[4-6, 11] This model involves annealing of a TAR DNA hairpin to a complementary TAR RNA hairpin, mimicking a key step in reverse transcription.[2, 12, 13]

Fluorescence resonance energy transfer SMS (SMFRET) has been extensively used to study two main aspects of annealing: (i) the kinetics of annealing under various conditions, and (ii) the secondary structure of nucleic acid intermediates in the annealing mechanism. In the recently proposed mechanism for the annealing of complementary TAR hairpins (eqns i.-iii. in Figure 4.1), the chaperone function of NC arises from two processes. First, NC partially melts the closed "C" secondary structure of TAR hairpins producing a partially open "Y" structure containing single-stranded regions that promote

the nucleation of annealing.[11, 14] In the presence of NC (and low Mg^{2+} concentration), "Y" is more stable than "C", due to preferential interaction of NC with the single strand regions of "Y". Second, NC increases the rate of the nucleation of annealing by weakly associating the two reacting hairpins in an encounter complex. [4, 6]

The major focus of this paper is the development of a molecular level understanding of the mode of action of model nucleic acid binding molecules, such as arginine and argininamide, for inhibiting annealing processes, especially minus strand transfer. Both molecules have been shown to exhibit specific binding affinity for the uracil bulge region in the TAR RNA hairpin.[15-18] The binding specificity between TAR RNA and the argininamide analogs is of interest because of the role that Tat, HIV's transcriptional transactivator, plays in binding to TAR RNA through its arginine residues at crucial steps during the viral life cycle.[19, 20] It has been proposed that the next generation of inhibitors may be developed from compounds that can bind TAR RNA similarly to the active native structure and induce structural conformations that inhibit functionality.[21] Such inhibition may involve a decrease in hairpin flexibility that is associated with argininamide binding.[22] To date, studies have demonstrated conclusively that argininamide binds to TAR RNA, [15-18] to TAR RNA/DNA composites,[23] and to various other DNA loop structures.[24, 25] Thus, it is important to obtain a clearer understanding of the role such inhibitors may play in the overall process of minus-strand transfer. No detailed analysis has been performed on either argininamide's specificity for the complementary TAR DNA hairpin, nor its efficiency at inhibiting minus-strand
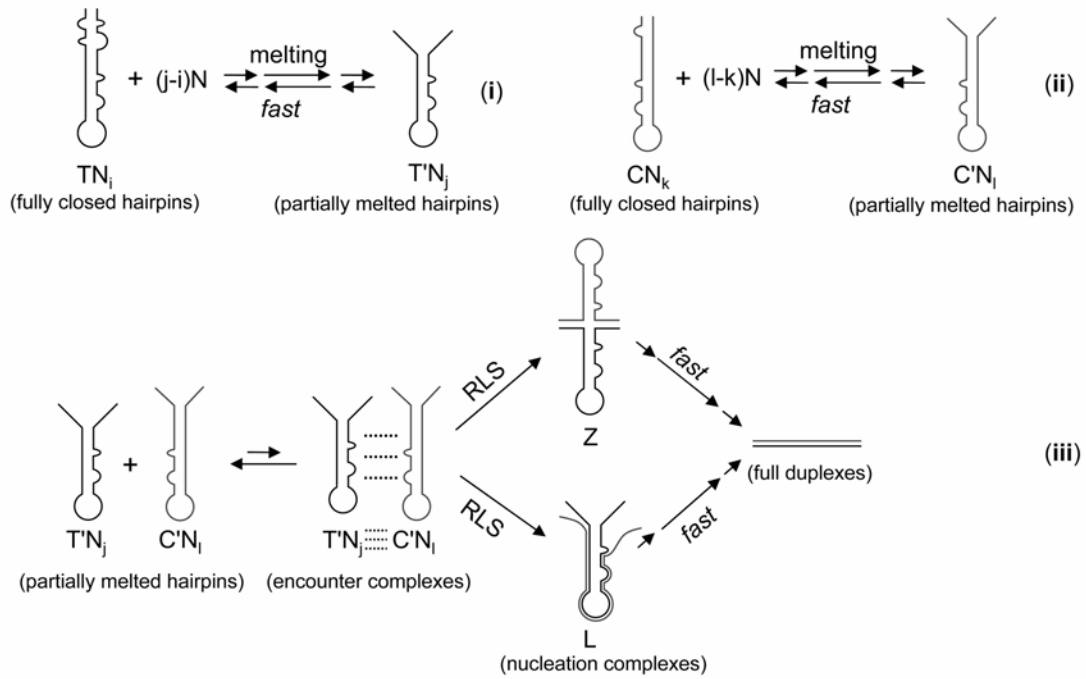
Figure 4.1 The hypothetical kinetic scheme of TAR DNA annealing to its complement chaperoned by HIV-1 NC. Here, T denotes TAR DNA, and C denotes complementary cTAR DNA or TAR RNA. The term N denotes the nucleocapsid protein, NC. In this scheme, N bound to T and C leads to a partially melted structure, namely the "Y" form of T (T') and C(C'). The subscripts, i, j, k and l are used to describe the number of NCs bound to nucleotides. Two partially melted hairpins form an encounter complex that leads to the formation of nucleation complexes. The annealing can go through either zipper nucleation or loop nucleation, thereby forming zipper nucleation complexes (Z) or loop nucleation complexes (L) that lead to the formation of fully annealed duplexes.

transfer. Such analyses, of even model binding agents, are important to the development of effective inhibitors.

Herein we use SMFRET to study how argininamide, as compared to the 'simple' divalent cation Mg^{2+}, modulates the chaperone activity of NC by examining its effect on the annealing kinetics of TAR DNA with complementary oligonucleotides, *in vitro*. Our goal is to obtain detailed information on the molecular scale action of HIV-1 inhibitors for minus-strand transfer, which could aid in the development of better inhibitors. The single molecule studies reveal that argininamide measurably suppresses the NC-induced partial melting of TAR. Previous studies with short targeted oligonucleotides suggest that annealing can be nucleated in either the hairpin loop region of TAR producing a so-called loop "L" nucleation complex, or in the terminal region of TAR, producing a so-called zipper "Z" nucleation complex.[4, 5] It is appropriate to use DNA analogs for these types of reactions because we have recently offered evidence that the reaction results are consistent with the DNA-RNA annealing reactions associated with minus-strand transfer.[6,5] By using short, targeted oligonucleotides we examine the effects of argininamide on each of these two previously identified sites for the nucleation of annealing. Argininamide is reported herein to inhibit NC-chaperoned zipper and loop type annealing with higher inhibition of the latter. The effects of argininamide on annealing are compared to the effects of varying concentrations of Mg^{2+}, which also is found to exhibit pathway-specific inhibition.

Experimental Section

Sample Preparation. Purified DNA sequences (see Figure 4.2) containing the appropriate functionalization (i.e. biotin for hairpin immobilization, and donor (Cy3) and acceptor dye (Cy5)) were acquired from TriLink Biotechnologies (San Diego, CA). The HIV-1 NC protein for these experiments was prepared by solid-phase synthesis or by *E. coli* expression, both described previously.[4, 11] All reagents were molecular biology and/or HPLC grade.

Clean coverslips were treated with Vectabond/acetone 1% w/v solutions (Vector Laboratories, Burlingame, CA) for five minutes. The coverslips were then rinsed with H_2O and dried under N_2. Each coverslip was pegylated and biotinylated, after which a reaction chamber with inlet and outlet ports (Nanoport[TM], Upchurch Scientific, Oak Harbor, WA) was glued on top. The chambers were rinsed with distilled deionized water (50 µL) following each incubation step. The details of this process are described in earlier work.[4]

	Primary Structure	
D/A-TAR DNA	5'-Cy3-TGGGTTCCCTAGTTAGCCAGAGAGCTCT(biotin)CAGGCTCAGATCTGGTCTAACCAGAGAGACCCTTTT-Cy5-3'	
D-TAR DNA	5'-Cy3-TGGGTTCCCTAGTTAGCCAGAGAGCTCT(biotin)CAGGCTCAGATCTGGTCTAACCAGAGAGACCCTTTT-3'	
inverted D-TAR DNA	5'-Cy3-TGGGTTCCCTAGTTAGCCAGAGAGCTCTCAGGCTCAGATCTGGTCTAACCAGAGAGACCCTTTT(biotin)-3'	
A-zipper mimic, Z	5'-GGGTCTCTCTGGCTAGGGAACCCTTTT-Cy5-3'	
A-loop mimic, L	5'-AGATCTGAGCCTGAGAGCTCTCTT-Cy5-3'	

Zipper Region

Loop Region

5' 3'

Figure 4.2 (*left*) The primary sequences of the DNA oligonucleotides used in the present work are listed, along with the Cy3/Cy5 and biotin functionalization of each species. (*right*) The predicted secondary structure of TAR DNA along with the regions of complementarity for each of the oligonucleotides (shaded regions) are illustrated.

Each chamber was treated with streptavidin (Molecular Probes, Eugene, OR; 0.2 mg/mL in 25 mM HEPES buffer) and subsequently with the appropriately labeled biotinylated TAR DNA. Standard D-TAR DNA and inverted D-TAR DNA were used interchangeably, depending on the type of reaction being performed. Past experiments have demonstrated a small (but measureable) effect of loop annealing rates depending on the biotin location in the loop region.[4] Reactant solutions, which were delivered by syringe pumps[4, 11] contained HEPES buffer (25 mM), $MgCl_2$(various concentrations as noted in the text), NaCl (40 mM), and an oxygen scavenger system (2-mercaptoethanol 1% v/v, β-D(+)glucose 3% w/v (Sigma-Aldrich, St. Louis, MO), glucose oxidase 0.1 mg/mL) and catalase 0.02 mg/mL (all Roche Applied Science, Hague Road, IN)).[26, 27] Additional reactants were 400 nM NC and L-argininamide•2HCl (Sigma-Aldrich, St. Loius, MO) in the buffered reactant solution described above. The Na^+ concentration for all reactions was 40 mM. The Mg^{2+} concentration was varied for each experiment and the relevant concentrations are noted in the figure captions.

Data Collection. The home-built sample scanning confocal optical/data collection system used in these measurements is based on a Zeiss inverted microscope.[11] The sample flow-cell is scanned by a Queensgate X,Y scanning stage (NPW-XY-100A, Queensgate, Torquay, UK). A high numerical aperture, oil immersion microscope objective (Zeiss Fluar, 100x, NA 1.3) was used to focus the excitation light (514 nm) and collect the emission. The donor and acceptor fluorescence were separated by a dichroic beam splitter (Chroma 630 DCXR, Chroma Tech., Vermont) into two beams, and each

detected by an avalanche photodiodes (APD) (Perkin Elmer Optoelectronics SPCM-AQR-15, Vaudreuil, Quebec, Canada).

Two modes were used to collect SMFRET data in this paper, the confocal kinetic mode and the individual trajectory mode. In the confocal kinetic mode, scanning confocal donor and acceptor images were acquired at several times, t, during the course of annealing reactions[4]. The annealing reactions were initiated by beginning the flow of a freshly prepared solution of NC plus either Z or L oligonucleotides into the flowing sample cell, which already contained a cover-slip coated with immobilized TAR hairpins. In the argininamide reactions, argininamide was pre-incubated with the appropriate buffer/salt solution and/or the appropriate complementary DNA sequence, and flowed into the reaction chamber for measurement. Donor and acceptor emission intensities were calculated from the individual fluorescence spots (due to individual immobilized dye-labeled hairpins) in the confocal images at various elapsed times, t, from the initiation of the Z or L oligonucleotide flow. The time spacing between images was typically hundreds of seconds for longer reactions, but was reduced to 25 seconds for increased resolution in faster reactions. In-house software was written to find molecules, correct for drift, photobleaching, blinking and crosstalk, and calculate SMFRET efficiency, as described below, for each molecule.

In the individual trajectory mode, the microscope was positioned to focus on an individual immobilized hairpin, and time, t, trajectories of donor and acceptor emission intensities were recorded with 1 ms time resolution for several seconds. The procedure was repeated for hundreds of molecules for each sample.

For both individual trajectory mode and kinetic mode, the corrected donor and acceptor intensities, $I_D(t)$ and $I_A(t)$, respectively, were used to calculate directly the time trajectory of the apparent FRET efficiency, $E_A(t)$, as in:

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \tag{1}$$

$E_A(t)$ is related to the actual FRET efficiency, $E_{FRET}(t)$, by the inclusion of the dye quantum efficiencies, ϕ_i, and detector quantum efficiencies, η_i as in:

$$E_{FRET}(t) = \frac{I_A(t)}{I_A(t) + I_D(t) \dfrac{\phi_A \eta_A}{\phi_D \eta_D}} \tag{2}$$

In the case of the current experimental setup, it was determined that $E_A(t) \sim E_{FRET}(t)$. The collected donor and acceptor signals were corrected for background emission/noise and donor/acceptor cross talk due to overlapping emission as previously described.[11]

For the *individual trajectory mode* only, single molecule cross correlations of $I_D(t)$ vs. $I_A(t)$ were calculated by Equation 3.

$$C(\tau) = \frac{\langle \delta I_D(t) \delta I_A(t+\tau) \rangle}{\langle I_D \rangle \langle I_A \rangle} = \frac{\langle I_D(t) I_A(t+\tau) \rangle}{\langle I_D \rangle \langle I_A \rangle} - 1 \tag{3}$$

E_A autocorrelation analysis was acquired using the single molecule cross correlation curves according to Equation 4, as described in detail in a previous study.[11] The measured autocorrelation amplitude will be referred to as $E_A AC$.

$$E_A Autocorrelation(\tau) = C(\tau)\left(-\frac{\langle I_D \rangle}{\langle I_A \rangle} \right) \tag{4}$$

Results and Discussion

Basic annealing kinetics. The kinetics of NC-induced annealing of donor labeled TAR hairpins (D-TAR) with each complementary DNA oligonucleotide, i.e. A-zipper mimic (Z) and A-loop mimic (L), were investigated by time resolved SMFRET. The primary structures of the various oligonucleotides are shown in Figure 4.2. The "D" or "A" preceding each sequence term implies a donor or a acceptor dye, respectively. SMFRET time trajectories for ~100 single TAR DNA hairpins annealing with each olignucleotide (in the absence or presence of argininamide) can be measured by the confocal kinetic mode, as described in the data collection section, above. Examples of both Z and L annealing (in the absence of argininamide) are shown in Figure 4.3 A and B, respectively. By examining the time trajectories for each type of reaction, one can obtain a statistical distribution of reaction times (observed in Figure 4.3A as a distribution of on times required for different molecules to switch from reactant, with E_A ~0, to product, with E_A ~ 0.96 for the D-TAR-Z complex) that yield an experimentally determined rate constant. One can also observe directly the equilibrium process for each two-state reaction (as exhibited by discrete jumps between the reactant un-annealed hairpin, with E_A ~0, and an annealed D-TAR/Z complex, with E_A~ 0.96.[4] Analogous two-state annealing was observed for D-TAR/L annealing (Figure 4.3B), although the annealed form in this case has a lower FRET value (E_A ~0.8) due to a longer D/A distance. Although the temporal resolution in the confocal kinetic mode is lower than with the individual trajectory mode, by comparing the two types of annealing in this way we can observe a statistically significant number of molecules, and simultaneously

confirm that, indeed, Z-type annealing is both faster and more reversible than L-type annealing. For the faster Z-type annealing, collection times were decreased to improve temporal resolution.

The two-state nature of the annealing process is especially apparent by analyzing an ensemble of SMFRET trajectories in the form of SMFRET histograms, as shown in Figures 4.4A-C for D-TAR molecules annealing to L at various reaction times in the absence of argininamide. These data clearly show two distinct peaks, which are assigned to the unannealed and annealed forms (see the schematic structures at the top of Figure 4.4). There is excellent agreement between the observed histograms and simulated histograms (blue curves) generated from a two-state, irreversible, first-order model at $t = 0$, $t = \tau$, and $t = \infty$ (Figure 4.4 A, B, and C, respectively). The small residual differences (green) between each simulated histogram and the corresponding experimental set are likely due to slight imperfections in the model, which is described in detail in the Supplementary Material. This type of comparison is especially important at intermediate reaction times, such as a t ~τ, where subtle deviations from two-state behavior might be expected in the experimental system. Deviation from two-state behavior is, however, observed neither for TAR DNA/L annealing, nor for the similar reaction in the presence of argininamide (Figures 4.4D-F), nor for TAR DNA/Z annealing (data not shown) in the presence or absence of argininamide.
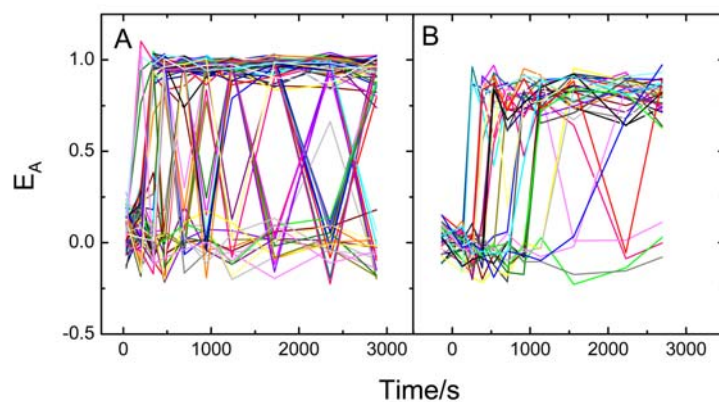
Figure 4.3 Single molecule E_A trajectories of ~ 100 molecules during the annealing of (A) 10 nM A-zipper DNA or (B) 10 nM A-loop DNA to immobilized D-TAR at 1 mM Mg^{2+} and 440 nM NC, collected by the confocal kinetic mode.
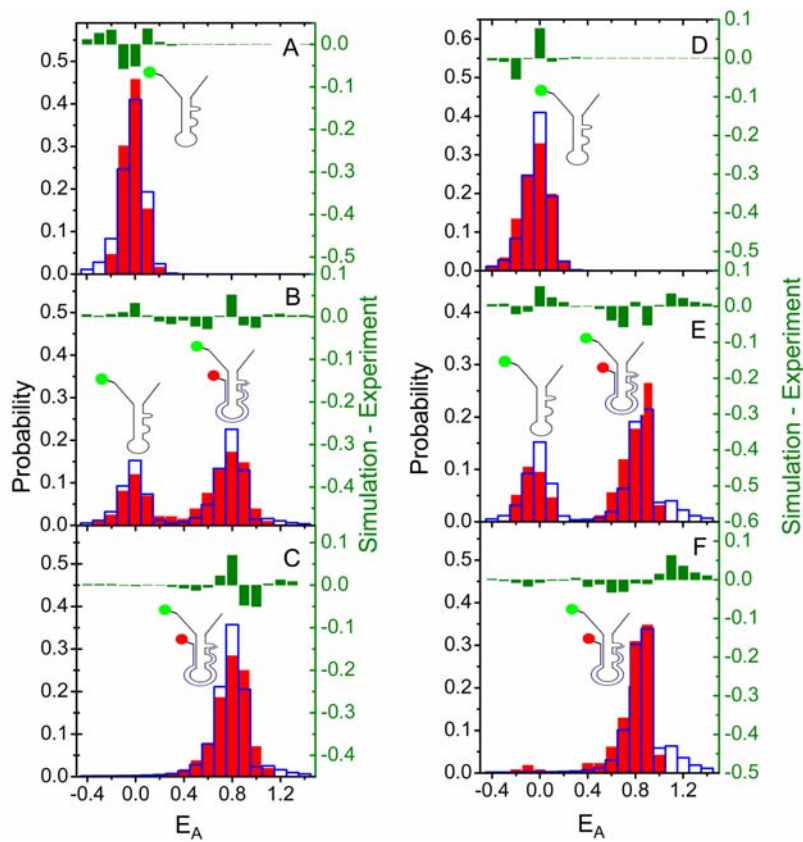
Figure 4.4 (Left) Experimental (red) and simulated (blue) SMFRET histograms for the reaction of immobilized D-TAR molecules with A-loop mimic (10 nM) at 1 mM Mg^{2+} (red) in the absence (A-C) and presence (D-F) of 2 mM argininamide at $t = 0$ (A, C), $t = \tau$ (B, E), and $t = \infty$ (C, F). The difference between the simulated and experimental histograms is shown in green.

The dynamics of L and Z type annealing have been shown to depend on many factors such as oligomer concentration, NC concentration, and the ionic strength of Mg^{2+} and to a lower extent, Na^+.[4, 11, 14] The complex role of Mg^{2+} will be discussed in more detail below, but one can examine the role of argininamide in pseudo-first-order conditions by maintaining a constant, high concentration of all species relative to the immobilized TAR DNA hairpin, and varying only the argininamide concentration. Figure 4.5 shows that the observed annealing kinetics are well-fit by a simple first-order treatment. The percent reaction in the figure has been calculated directly from $E_A(t)$ data using a E_A threshold of 0.4 to distinguish annealed and unannealed hairpins. The observed simple-first order behavior is, furthermore, consistent with a pseudo first-order kinetic scheme,

$$TAR + L/Z \xrightleftharpoons{[NC]} P \qquad (5)$$

resulting from the large excess concentration of Z or L hairpins and the high NC concentration, which ensures saturated NC binding. The annealing kinetics are approximately irreversible, under the relatively high Mg^{2+} concentration used in this study, but are reversible at lower Mg^{2+} concentration, as described earlier.[4] As previously demonstrated,[4] the NC-chaperoned loop-type annealing occurs at a slower rate than zipper-type annealing.
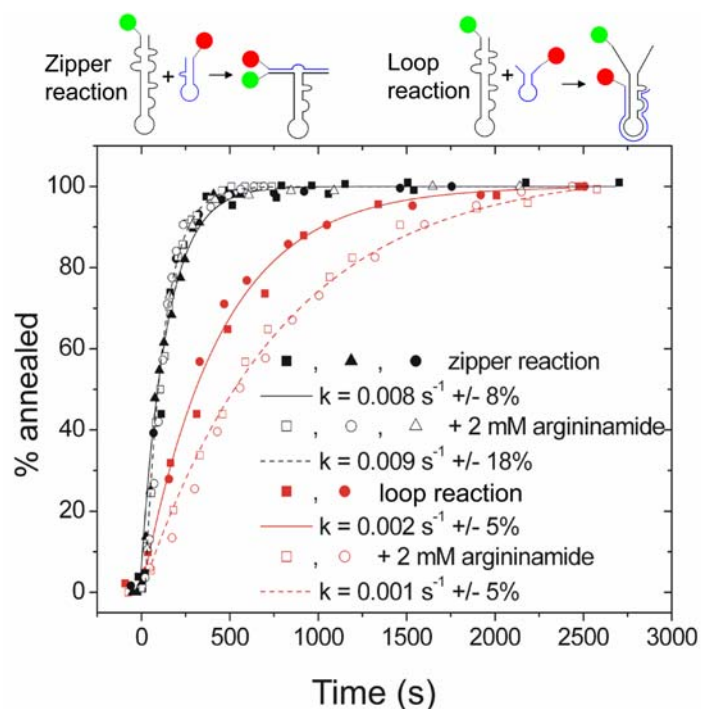
Figure 4.5 The effects of argininamide (2 mM) on NC-chaperoned loop-type (*red*) and zipper-type (*black*) annealing kinetics (with 10 nM of each oligomer in solution) are compared in the presence of 1 mM Mg^{2+} and 400 nM NC. The curves associated with each type of reaction correspond to a single exponential fit with the rates indicated in the inset. The fits were achieved using three separate experimental time trajectories for each zipper reaction and two separate experimental trajectories for each loop reaction, as indicated by the different data symbols in the figure. The data/curves have been normalized, for presentation.

As summarized in Figure 4.5, the data from 10 separate experiments show that the Z-type annealing rate is unaffected, whereas the L-type reaction rate is decreased by 50%, in the presence of 2 mM argininamide. It is important to note that in this example, the argininamide is present at twice the Mg^{2+} concentration, which is in itself high enough to influence strongly the equilibrium stability of the double-stranded DNA products or reactants (see Ref. 14 for a detailed discussion). Each reactant oligomer is present in nanomolar amounts, and the immobilized TAR DNA concentration is immeasurably lower. Table 1 compares second order rate constants for Z and L annealing reactions at several different argininamide concentrations, including a limiting case performed at 10 mM that will be discussed below. Additional data from experiments performed at 0.2 mM argininamide (at which concentration Mg^{2+} has been shown to affect the annealing dynamics), are not included because they are indistinguishable from the control. The results can be summarized as follows: it is possible to observe specificity of argininamide in inhibiting L-type annealing, as highlighted by the 2 mM data; at very high concentrations both pathways are inhibited, most likely due to high ionic strength; and meaningful inhibition of either reaction, regardless of its molecular origin, can only be achieved at a prohibitively (and toxically) high argininamide concentration. These results can be contrasted with the Mg^{2+} dependence of each pathway. The kinetic curves shown in Figure 4.5 were obtained at 1 mM Mg^{2+}. When the reactions are performed at 2 mM Mg^{2+},[4] both zipper and loop annealing rates decrease, with a more dramatic effect observed for the loop rate. Additional data on this topic is included in the Supplementary Material.

Loop vs. Zipper Annealing. We have observed that argininamide can inhibit the rate of NC-chaperoned annealing. The inhibitory effect of argininamide is, however, greater for loop-type annealing than for zipper-type annealing, see Figure 4.5 and Table 1. For argininamide concentrations higher than 2 mM, the annealing rates were too slow to measure accurately by direct kinetic measurements of the type in Figure 4.5 due to dye photobleaching and sample drift. Fortunately, the rates of the zipper vs. loop modes could be estimated from single time-point (extent of reaction) determinations (assuming first-order kinetics) using the relative concentration of annealed and unannealed hairpins at a chosen time after initiating the reaction. The relative concentrations were determined from E_A histograms of the type shown in Figure 4.6 ,which were calculated from a set of E_A trajectories, recorded in *individual trajectory mode,* in a short time interval ($\sim$ 30 min) after the annealing reaction was initiated.

Table 4.1 The apparent second-order rate constant k_a for the NC chaperoned annealing of TAR DNA to its zipper/loop complementary oligomers at various argininamide concentrations.[a]

[Arg] (10⁻³ M)	Zipper (10⁵s⁻¹M⁻¹)	Loop (10⁵s⁻¹M⁻¹)
0	8 ± 1^b	2.0 ± 0.1^b
1	7^c	2^c
2	9 ± 2^b	1.00 ± 0.05^b
10	0.03^d	0.002^d

[a] The NC and Mg^{2+} concentrations were 400 nM and 1 mM, respectively.

[b] Values were determined from the single molecule kinetic measurements shown in Figure 4.5

[c] Error was not estimated.

[d] Due to the increasing impact of artifacts, such as photobleaching, on single molecule counting when the rate is slow, especially at high argininamide concentration, these values were estimated from the pseudo first-order rate equation $\ln \frac{R}{R_0} = -kt$. The ratio of the number of unreacted molecules (R) after time t to the number of molecules before reaction started(R_0) was determined from the ratio of two peak areas in the E_A histogram shown in Figure 4.6, and it was calculated to be 0.726 for zipper annealing and 0.976 for loop annealing. The t was roughly estimated to be 3 hours.

At negative times, i.e. before the non-immobilized hairpin was flowed into the sample cell, the E_A histogram only exhibited a single peak, i.e. due to unannealed D-TAR, with E_A ~ 0 (Figure 4.6A). Figure 6B portrays a histogram calculated from trajectories that were recorded 30 minutes after the initiation of D-TAR annealing with the loop oligonucleotide. The two peaks in the histogram are assigned to the TAR/loop annealed product (E_A ~ 0.8) and the unreacted D-TAR (E_A ~ 0) hairpins, resulting from the slow annealing kinetics. The E_A histogram for the analogous reaction with the A-zipper mimic collected after ~ 30 minutes reveals 100% formation of the TAR/zipper annealed product, due to the faster annealing rate for this type of nucleation (Figure 4.6E).

As mentioned above, the inhibition of annealing is considerably more effective for loop type annealing (Figure 4.6D) than zipper type annealing (Figure 4.6C). The loop annealing rate is reduced by a factor of 1000 by 10 mM argininamide, while the zipper annealing rate is only reduced by a factor of ~300 at this saturating argininamide condition (Table 4.1). This effect is even more apparent at 2 mM argininamide, where the zipper annealing rates are not significantly different from the argininamide free results, but the loop annealing rate is reduced 2 fold in its presence. Further evidence that argininamide inhibits loop-type annealing more effectively than zipper-type annealing is shown in Figure 4.7, which portrays the observed ratio of rates for zipper vs. loop annealing as a function of argininamide concentration.
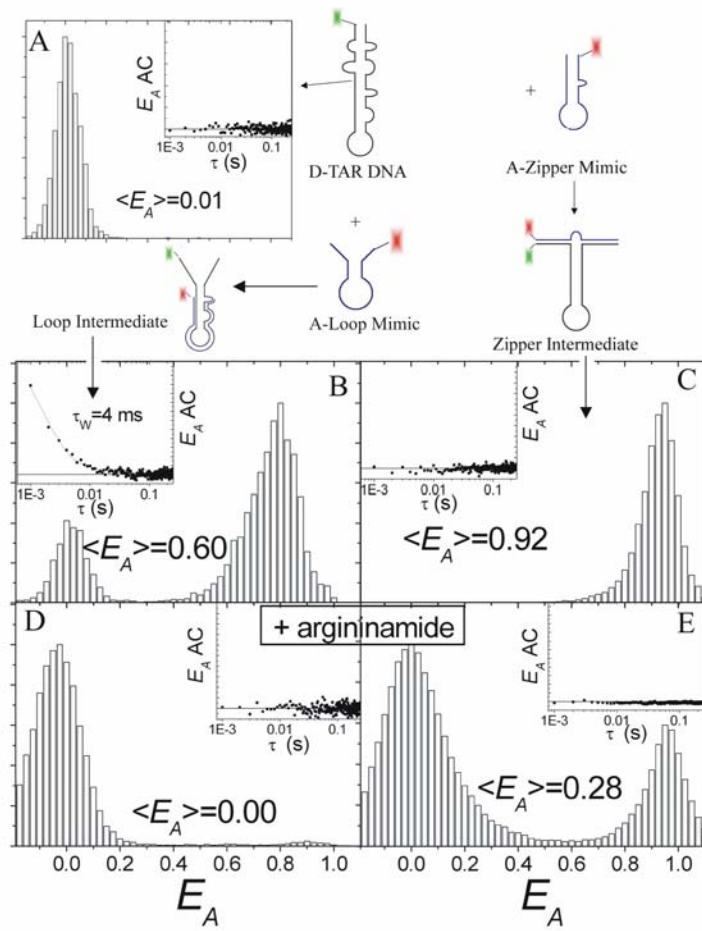
Figure 4.6 (*A*) The ensemble E_A histogram for D-TAR DNA is shown. (*B,C*) The ensemble E_A histograms of the loop (B) and zipper (C) reactions alone and in the presence of 10 mM argininamide (D, E) are presented with the respective E_A autocorrelations (*insets*).
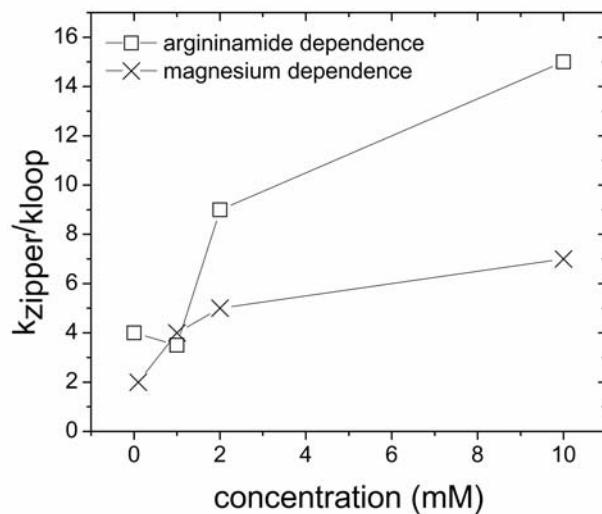
Figure 4.7 A comparison of the argininamide and magnesium dependence of the relative reaction rates for zipper vs. loop annealing is shown. The relative rates at 10 mM argininamide were calculated from the respective peak areas of the products vs. reactants. The lines are included as a guide only.

TAR DNA structure. The effect of argininamide on NC's destabilization activity was investigated by recording the ensemble E_A histogram for immobilized double-labeled D/A-TAR DNA in the absence and presence of NC and argininamide. (These data were recorded without a complementary oligonucleotide present, since the goal here is to measure the intermolecular distribution of 3'/5' distances for TAR DNA, rather than the annealing kinetics.) As previously demonstrated, and reproduced in Figure 4.8 (top left), D/A-TAR DNA in the absence of NC exists exclusively in its relatively rigid closed "C" form[11]. The "C" form, which has the maximum number of possible Watson-Crick base pairing for this hairpin, is characterized by $<E_A>$ values of 0.95 – 0.96 (indicating a relatively short 3'/5' distance) and a negligibly small autocorrelation amplitude (indicating only small amplitude 3'/5' excursions).[14] The presence of argininamide has, as expected, no effect on the "C" form (Figure 4.8, top right) since binding with this ligand tends to increase the strength of Watson-Crick base pairing, thus favoring the "C" form.

As shown previously, when D/A-TAR DNA is in the presence of NC at concentrations greater than 100 nM , the secondary structure in the region near the two bulges at the top of the stem is disrupted, resulting in a partial population of the "Y" form, [11] and a resultant shift of $<E_A>$ to 0.86. The "Y" form is manifested in the E_A histogram as an additional peak at lower E_A than the peak assigned to C (Figure 4.8, bottom left). A second manifestation of NC induced partial melting is a large increase in the amplitude of the E_A autocorrelation amplitude, reflecting NC induced DNA secondary structure fluctuations, as described in previous studies.[11][14]
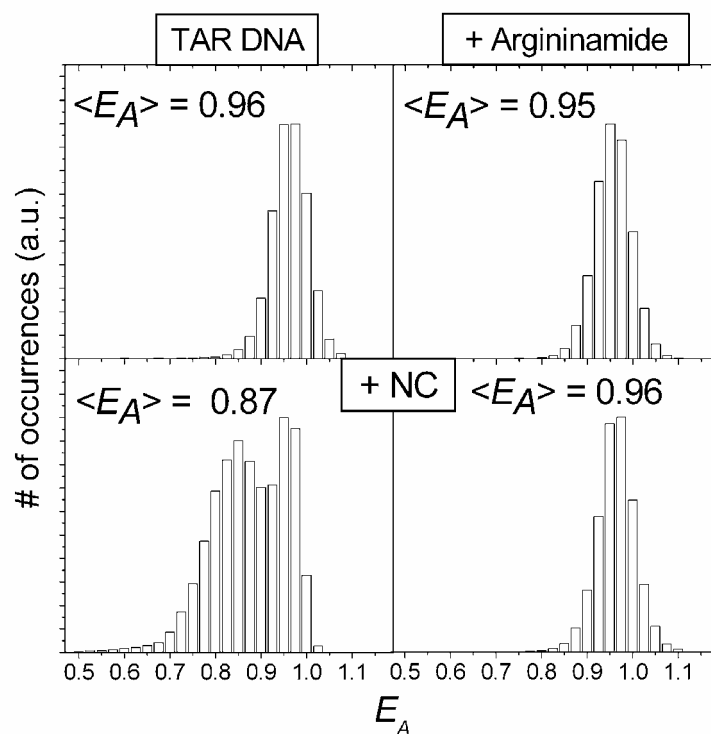
Figure 4.8 (*left*) E_A histograms (binning time=10 ms) for D/A-TAR DNA in buffer alone (*top*), and in the presence of 440 nM NC (*bottom*) are shown at 2 mM Mg^{2+}. (*right*) The results for the same experimental conditions as in the left-hand panels, but in the presence of 10 mM argininamide are compared.

The suppression of NC activity on TAR DNA secondary structure by argininamide varies from no suppression at sub milli-molar concentrations to nearly complete suppression at 10 nM argininamide (Figure 4.8, bottom right). For example $<E_A>$ of D/A-TAR DNA in the presence of NC and argininamide concentrations ranging from 0.2 to 10 mM were acquired at 2 mM Mg^{2+} (Figure 4.9). These results indicate that at low argininamide concentration (e.g. 0.2 mM), the effects on NC's ability to destabilize the secondary structure are negligible. Above 2 mM argininamide the E_A analysis shows that the average TAR DNA secondary structure is primarily in the "C" form. E_A autocorrelation analysis, however, reveals that secondary structure fluctuations are still measurable at all but the highest argininamide concentration (Figure 4.9). Interestingly, whereas the E_A autocorrelation amplitude decreased with addition of argininamide, the E_A autocorrelation lifetime remained constant (data not shown), indicating that although the argininamide induces an overall shift in the equilibrium away from the "Y" form of TAR DNA, when the hairpin occupies this state it exhibits the same complex dynamics reported in earlier studies to be important for annealing to occur.[11][14]

A mechanistic interpretation of the kinetic inhibitory effect of argininamide. We have previously hypothesized that the NC-induced formation of the Y secondary structure of TAR DNA provides a low activation energy pathway for annealing of TAR DNA with oligonucleotides. The correlation observed herein between effects of added argininamide on the equilibrium population (concentration) of the "Y" form and the annealing rates of TAR DNA with oligonucleotides is highly consistent with this hypothesis. Qualitatively, these effects can be understood in terms of the model in Figure

4.1. The addition of argininamide apparently increases the relative stability of duplex secondary structures, i.e. those with greater Watson-Crick base-pairing, thereby reducing the relative population of the partially melted oligonucleotides, e.g. "Y", which are key intermediates in the annealing mechanism.

Binding between DNA hairpins and arginine (an analogous ligand to argininamide) has been previously observed, although shown to be weaker than binding to the analogous RNA hairpin.[23] Other studies suggest that DNA/argininamide binding occurs through an alternate mechanism (i.e. through minor groove hydrogen bonding interactions with base edges rather than with phosphate groups).[24] The distribution of binding sites on TAR DNA in the annealing reaction is not well established, but must be extremely complex, since partly melted TAR DNA presents multiple thermally accessible secondary structures with different degrees of melting of the four internal bulges.[11] We hypothesize that the observed preferential inhibition of loop annealing may be due to a selective stabilization of this region by argininamide, although there is no direct evidence for this proposal.
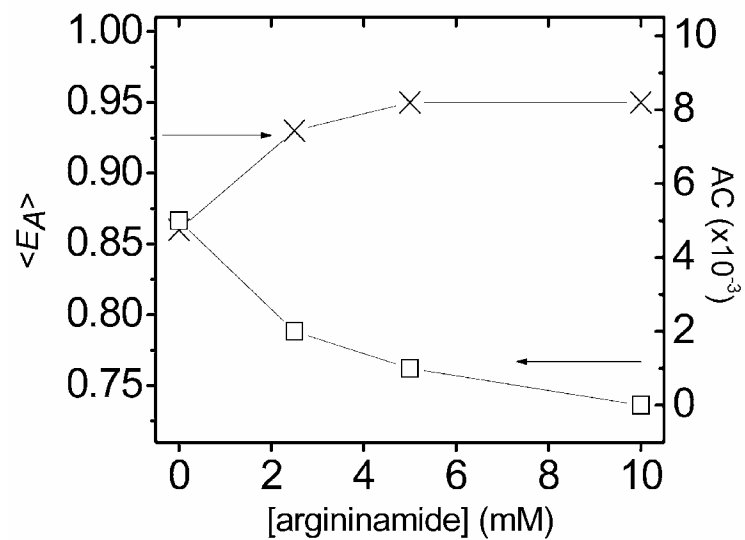
Figure 4.9 (*left*) The average E_A for D/A-TAR DNA in the presence of 440 nM NC and various concentrations of argininamide are compared. (*right*) The trend in average E_AAC at each concentration is shown.

It is important to note that the general effects of argininamide on the equilibrium structure of TAR DNA and on the annealing rate are analogous to those reported previously by us[4, 11] for Mg^{2+}, namely that they both decrease NC-induced TAR DNA fluctuations and slow down annealing kinetics with varying degrees of efficiency. Additionally, as shown in Figure 4.9, increased Mg^{2+} concentrations preferentially suppress loop-type annealing in a parallel fashion to argininamide. Furthermore, both argininamide and Mg^{2+} have been shown by others to decrease fluctuations in the TAR RNA hairpin [17, 28]. The various effects of Mg^{2+} may reflect both non-specific charge interactions as well as specific binding interactions such as those reported previously for Mg^{2+} binding to the hairpin-loop region of TAR RNA. [29] There does not exist, to our knowledge, similar structural characterizations by techniques such as NMR or x-ray analysis for the specific binding of Mg^{2+} or argininamide to the TAR DNA hairpin. Such analysis would be helpful to supplement what we have learned about the dynamics of strand transfer, namely that both ions exhibit an enhanced inhibitory effect for annealing that involves the hairpin loop region.

However, both charge-shielding and specific binding effects have been observed for Mg^{2+}, e.g. complex Mg^{2+}-dependent conformational changes were observed in an immobilized RNA three-way junction.[30] Thus, it is likely that both specific and non-specific outer-sphere charge shielding by Mg^{2+} and argininamide play a role in the annealing reaction, i.e. by suppressing the binding of NC to TAR DNA. NC is known to favor binding to single-stranded nucleic acids, which increases the thermodynamic stability of partially melted structures like "Y".[11, 14] A decrease in the binding constant of

101

NC to single-stranded regions due to charge shielding would decrease the annealing rate

and the secondary structure fluctuation of TAR DNA.

Conclusions

SMFRET was used to examine how argininamide, a model inhibitor for HIV-1 functionality, modulates the chaperone activity of the nucleocapsid protein (NC) in the minus-strand transfer step of HIV-1 reverse transcription, *in vitro*. Argininamide is observed to measurably inhibit the rate of NC-chaperoned annealing of TAR DNA with targeted oligonucleotides. The studies reveal that the argininamide inhibitory mechanism involves a shift of the secondary structure of TAR, away from the NC induced partially melted "Y" form, and toward the NC free "C" form. It was also found that the NC induced nucleation of annealing of TAR is more effectively inhibited at the hairpin loop region rather than the terminal stem region, suggesting molecular level interactions of argininamide and NC binding to key annealing intermediates.

Supporting Information

Simulation of Ensemble SMFRET Histograms

It was assumed that the intensity distributions in both channels are Gaussian distributions, which in the normalized form are $I_d = \dfrac{1}{\sigma_d \sqrt{2\pi}} e^{-\frac{(i_d' - \overline{I_d})^2}{2\sigma_d^2}}$ for the donor channel and $I_a = \dfrac{1}{\sigma_a \sqrt{2\pi}} e^{-\frac{(i_a' - \overline{I_a})^2}{2\sigma_a^2}}$ for the acceptor channel. The probability density of measuring a FRET efficiency value (E_{Obs}) for a given $\overline{I_a}$, $\overline{I_d}$, σ_a^2 and σ_d^2, denoted as $P(E_{Obs})$, is the joint probability density of simultaneously having donor intensity at i_d' and acceptor intensity at i_a', while i_d' and i_a' satisfy the criteria $i_a'/(i_d' + i_a') = E_{Obs}$. The i_a' is then dependant on i_d' and denoted as $i_{a,\,i_d'}'$. Thus, the $P(E_{obs})$ as a function of E_{Obs} should be expressed as an integral and then simplified as follows:

$$P(E_{obs}) = \int_{-\infty}^{\infty} \frac{1}{\sigma_d \sqrt{2\pi}} e^{-\frac{(i_d' - \overline{I_d})^2}{2\sigma_d^2}} \cdot \frac{1}{\sigma_a \sqrt{2\pi}} e^{-\frac{(i_{a,\,i_d'}' - \overline{I_a})^2}{2\sigma_a^2}}\, di_d'$$

$$= \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4\sigma_d^2 \cdot \sigma_a^2}[2\sigma_a^2 \overline{I_d}^2 + 2\sigma_d^2 \overline{I_a}^2 - \frac{(-4\sigma_a^2 \overline{I_d} - 4\sigma_d^2 \overline{I_a} \cdot \frac{E_{Obs}}{1-E_{Obs}})^2}{8\sigma_a^2 + 8\sigma_d^2 (\frac{E_{Obs}}{1-E_{Obs}})^2}]} \frac{1}{\sqrt{2\sigma_a^2 + 2\sigma_d^2 (\frac{E_{Obs}}{1-E_{Obs}})^2}}$$

To plot $P(E_{Obs})$ vs. E_{Obs}, one needs to know the average donor and acceptor intensity ($\overline{I_a}$, $\overline{I_d}$) and the two variances (σ_a^2 and σ_d^2). The $\overline{I_a}$ and $\overline{I_d}$ are determined by experiment. The variances are calculated as the sum of signal intensity and background

intensity, i.e., $\sigma_a^2 = \overline{I_a} + I_{background,a}$ and $\sigma_d^2 = \overline{I_d} + I_{background,d}$, when taking into account the broadening effect contributed from background intensities. As experiment conditions such as the total fluorescence intensity, etc., vary slightly from experiment to experiment, the ensemble histograms differ from the histograms of each individual experiment. To take into account that difference, the parameters are slightly adjusted to make the simulation fit the ensemble histogram obtained at t =0, ∞ (when the reaction has reached equilibrium), and then used to simulate the ensemble histogram at t = τ. The simulation was performed in Mathematica; after which the simulated data was binned with the same bin size as experiment and normalized by the total area. The parameters used in the simulations are listed in Table S1. It is important to note that the assumption of Gaussian behavior will break down when the intensity for either channel equals zero. In this instance, a more appropriate model would be a Poisson distribution. The observable associated with this simplification in the model is that the predicted FRET values would over (or under) estimate the population when, respectively, the acceptor or donor intensities approach zero. This can be observed in the simulated FRET histograms in for example, Figure 4A and F.

Magnesium vs. Argininamide dependence of Zipper and Loop Annealing

Argininamide was found to have a stronger influence on L-type annealing than on Z-type annealing, as discussed in the main body of text. In contrast both rates were influenced by changes in the Mg^{2+} concentration, (but Z is more strongly influenced). Further study of Mg^{2+} effects on the dynamics of each type of reaction is planned, but the following figure provides sample Mg^{2+} data, for contrast.

105

Table 4-S1 Parameters used in the simulation.

In the absence of Argininamide:

	$\overline{I_a}$ (ms^{-1})	$\overline{I_d}$ (ms^{-1})	σ_a^2 (ms^{-2})	σ_d^2 (ms^{-2})
$t = 0$	0	17	2	18
$t = \infty$	14.6	2.4	18.6	3.4

In the presence of Argininamide:

	$\overline{I_a}$ (ms^{-1})	$\overline{I_d}$ (ms^{-1})	σ_a^2 (ms^{-2})	σ_d^2 (ms^{-2})
$t = 0$	0	17	2	18
$t = \infty$	15.6	1.4	19.6	2.4

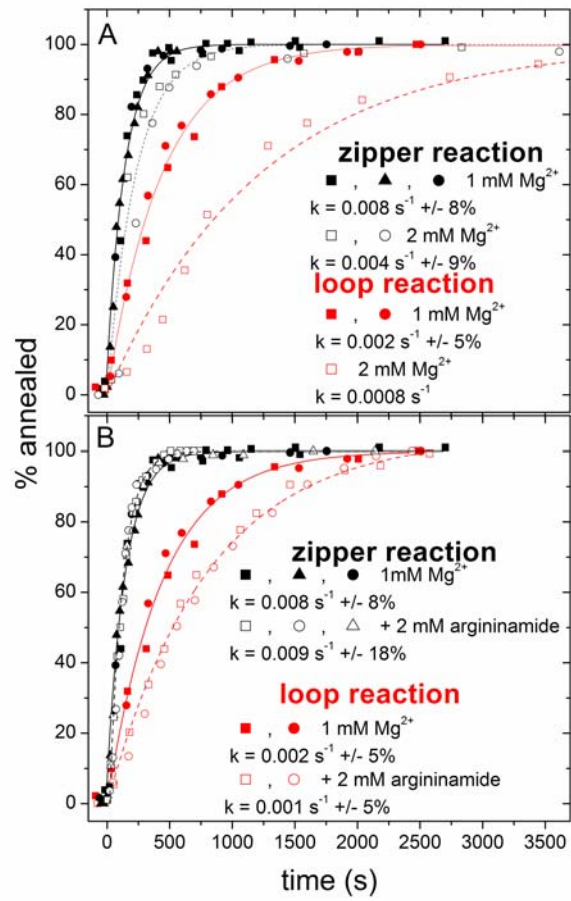Figure 4-S1. The effects of increased Mg^{2+} (A) and argininamide (B) are compared in normalized zipper/loop reactions. [NC] = 400 nM

Acknowledgements

References

(1)   Hong, M. K.; Harbron, E. J.; O'Connor, D. B.; Guo, J.; Barbara, P. F.; Levin, J. G.; Musier-Forsyth, K. *J. Mol. Biol.* **2003,** 325, 1.

(2)   Bernacchi, S.; Stoylov, S.; Piémont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mély, Y. *J Mol Biol* **2002,** 317, 385.

(3)   Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mély, Y. *J Mol Biol* **2003,** 328, 95.

(4)   Liu, H. W.; Cosa, G.; Landes, C. F.; Zeng, Y.; Kovaleski, B. J.; Mullen, D. G.; Barany, G.; Musier-Forsyth, K.; Barbara, P. F. *Biophys. J.* **2005,** 89, 3470.

(5)   Zeng, Y.; Liu, H. W.; Landes, C. F.; Kim, Y. J.; Ma, X.; Zhu, Y.; Musier-Forsyth, K.; Barbara, P. F. *Proc. Natl. Acad. Sci. USA* **2007,** in press.

(6)   Liu, H. W.; Zeng, Y.; Landes, C. F.; Kim, Y. J.; Zhu, Y.; Ma, X.; Vo, M.; Musier-Forsyth, K.; Barbara, P. F. *Proc. Natl. Acad. Sci. U S A* **2007,** 104, 5261.

(7)   Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G. *J. Virol.* **2000,** 74, 8980.

(8)   Guo, J.; Wu, T.; Kane, B. F.; Johnson, D. G.; Henderson, L. E.; Gorelick, R. J.; Levin, J. G. *J. Virol.* **2002,** 76, 4370.

(9)   Heilman-Miller, S. L.; Wu, T.; Levin, J. G. *J. Biol. Chem.* **2004**.

(10)   Levin, J. G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K. *Prog. Nucleic Acid Res. Mol. Biol.* **2005,** 80, 217.

(11)   Cosa, G.; Harbron, E. J.; Zeng, Y.; Liu, H. W.; O'Connor, D. B.; Eta-Hosokawa, C.; Musier-Forsyth, K.; Barbara, P. F. *Biophys. J.* **2004,** 87, 2759.

(12)   You, J. C.; McHenry, C. S. *J. Biol. Chem.* **1994,** 269, 31491.

(13)   Driscoll, M. D.; Hughes, S. H. *J. Virol.* **2000,** 74, 8785.

(14)   Cosa, G.; Zeng, Y.; Liu, H. W.; Landes, C. F.; Makarov, D. E.; Musier-Forsyth, K.; Barbara, P. F. *J. Phys. Chem. B* **2006,** 110, 2419.

(15)   Tan, R.; Frankel, A. D. *Biochem.* **1992,** 31, 10288.

(16)   Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992,** 257, 76.

(17)   Pitt, S. W.; Majumdar, A.; Serganov, A.; Patel, D. J.; Al-Hashimi, H. M. *J. Mol. Biol.* **2004,** 338, 7.

(18)   Dayie, K. T.; Brodsky, A. S.; Williamson, J. R. *J. Mol. Biol.* **2002,** 317, 263.

(19)   Berkhout, B.; Silverman, R. H.; Jeang, K. T. *Cell* **1989,** 59, 273.

(20)   Brigati, C.; Giacca, M.; Noonan, D. M.; Albini, A. *FEMS Microbio. Let.* **2003,** 220, 57.

(21)   Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I. H.; Karn, J.; Lentzen, G.; Drysdale, M.; Bower, J.; Potter, A. J.; Starkey, I. A.; Swarbrick, T.; Aboul-ela, F. *J. Mol. Biol.* **2004,** 336, 343.

(22)   Tinoco JR, I.; Collin, D.; Li, P. T. X. *Biochem. Soc. Trans* **2004,** 32, 757.

(23)   Landt, S. G.; Tipton, A. R.; Frankel, A. D. *Biochem.* **2005,** 44, 6547.

(24)   Robertson, S. A.; Harada, K.; Frankel, A. D. *Biochem.* **2000,** 39, 946.

(25)   Lin, C. H.; Wang, W.; Jones, R. A.; Patel, D. J. *Chem. & Biol.* **1998,** 5, 555.

(26)   Ha, T. *Curr. Opin. Struct. Biol.* **2001,** 11, 287.

(27)   Harada, Y.; Sakurada, K.; Aoki, T.; Thomas, D. D.; Yanagida, T. *J. Mol. Biol.*
        1990, 216, 49.

(28)   Al-Hashimi, H. M.; Pitt, S. W.; Majumdar, A.; Xu, W.; Patel, D. J. *J. Mol. Biol.*
        2003, 329, 867.

(29)   Olejniczak, M.; Gdaniec, Z.; Fischer, A.; Grabarkiewicz, T.; Bielecki, L.;
        Adamiak, R. W. *Nucleic Acids Res.* **2002,** 30, 4241.

(30)   Kim, H. D.; Nienhaus, G. U.; Ha, T.; Orr, J. W.; Williamson, J. R.; Chu, S. *Proc.
        Natl. Acad. Sci. U S A* **2002,** 99, 4284.

Chapter 5: Instrumentation for Single-Molecule FRET Detection

Introduction

In this chapter, instrumentation and software programming for the purpose of single-molecule fluorescence detection are discussed. Three microscopes have been developed during the dissertation research. The one-color excitation confocal microscope with two detection channels serves as a basic microscope to measure the single-molecule FRET. This microscope has been used to collect single-molecule donor/acceptor intensity trajectories (i.e. the transient mode) and to scan images containing over hundred molecules in the same region along annealing reaction (i.e. kinetic mode). This microscope has also been applied to Fluorescence Correlation Spectroscopy studies (FCS), by replacing the dichroic mirror before donor/acceptor channel with a 50/50 beam splitter and using the magnification ring to confine detection volume. The two-color alternative excitation microscope is developed upon this basis to measure the donor/acceptor fluorescence intensity generated by FRET, as well as the acceptor intensity generated by direct excitation with red laser. The strategy and instrument design of two-color microscope are discussed at technical level. Lastly, a brief discussion on a wide-field microscope is introduced. Evaluation of the wide-field microscope shows that even it has worse signal-to-background ratio than confocal microscope, the signal quality is good enough for single-molecule detection. The wide-field microscope, which is ideal for kinetic measurement, is limited by the photobleaching effect, however, and not suitable for slow kinetics.

Basic Setup

Most of the dissertation researches are carried out on a Zeiss confocal inverted fluorescence microscope, which is optimized for dynamical spectroscopic measurements with high sensitivity. The instrument utilizes epi-excitation at the top surface of the bottom cover-slip of the sample cell, where the NC/nucleic acid complexes will be bound for experiments. To collect single-molecule fluorescence intensity trajectories or images, the magnification after microscope is set to 1x to increase signal, give total magnification of 100x.

However, the magnification at the detectors for the microscope can be easily switched to 2.5x by a selectable lens in order to achieve confocal detection to confine the detection volume allowing for higher detection efficiency such as Fluorescence Correlation Spectroscopy (FCS). The fluorescence spots after microscope is ~ 90 μm in diameter, close to the active detection area of the detector (130 μm), which serves as a pinhole to effectively reject the out-of-focus fluorescence.

A continuous wave excitation from an Argon ion (514 nm) or He-Ne laser (633 nm) is directed by a dichroic mirror (Z514/633 rpc, Chroma, Rockingham, VT) into an objective (Zeiss Fluar, 100x, N.A. 1.3) and focused on coverslip surface (for single molecule imaing) or into the solution 20 μm above the coverslip surface (for FCS measurements). The collected fluorescence signal is filtered from the scattered light by a holographic notch filter (Kaiser Optical Systems, Ann Arbor, MI), refocused by a 1x relay lens and separated into two avalanche photodiodes (APD) detectors (SPCM-AQR-15, Perkin Elmer Optoelectronics, Vaudreuil, Quebec, Canada) by a 650 nm dichroic

mirror (or 50/50 beam splitter (Melles Griot, Rochester, NY) for FCS studies). The TTL

signals from APDs are distributed by a fan-out buffer (Pulse Research Labs, Torrance,

CA) into a real-time digital correlator (ALV-5000, ALV, Langen, Germany) and a

counter board (Becker & Hickl PMS-400, Berlin, Germany), which allowed for

simultaneous acquisition of fluorescence autocorrelation curves and fluorescence

intensity trajectories.

Intensity Trajectory Data Analysis Program SMS

The intensity trajectory of each single molecule was analyzed by a home-built

Matlab program called SMS with Graphical User Interface (GUI). The whole program

package contains two parts with total 132 files of hundreds of pages long (not shown due

to the size limitation of this dissertation). The first part of the program creates an

experimental data record and reads in all the transients. Then, one can choose the part of

transient to analyze one-by-one with mouse click on GUI. After each analysis, the

autocorrelation, the FRET histogram, the intensity distribution, etc. for that transient will

be shown for review, and then the next transient will appear for analysis. Once all the

transients are done, user goes to the second stage of the program to get sub-ensemble

results according certain tunable parameters. The beauty of SMS lies in that it not only

shows statistic information of a population of molecules, but also obtain the histograms

for a sub-population of molecules conveniently by using all kinds of thresholds such as
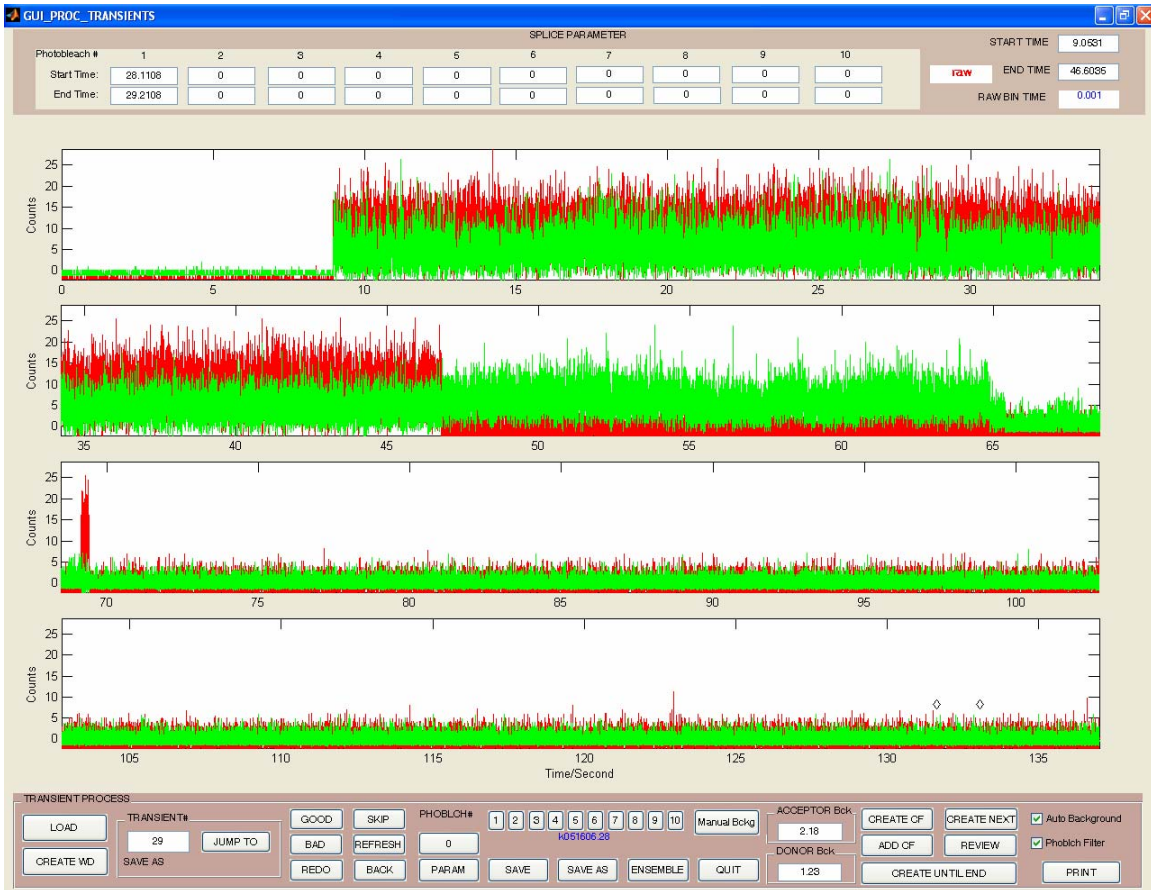
FRET, intensity and autocorrelation amplitude.

Figure 5.1 The main SMS GUI interface for loading data and performing analysis in an interactive manner. User will select the start and end of trajectory by clicking on the graph, then choosing the background correction.

115
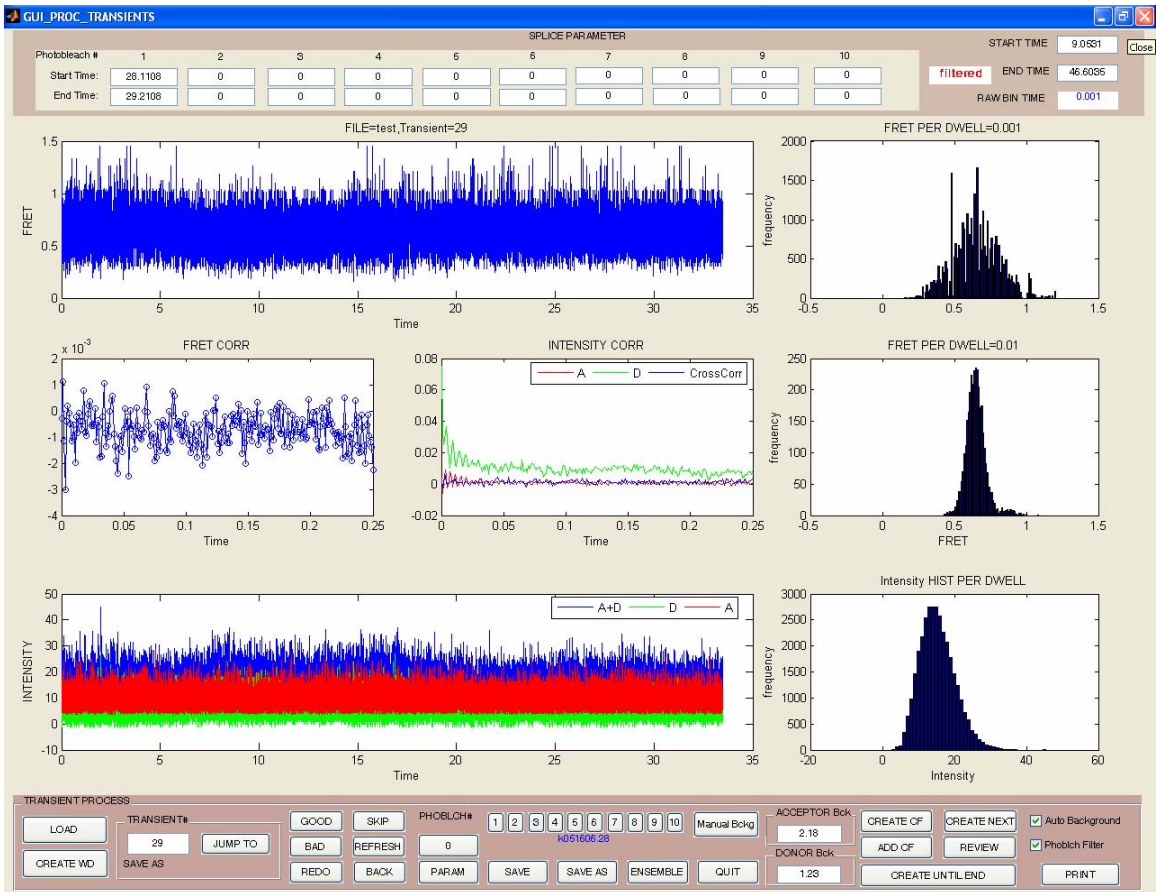
Figure 5.2 The SMS GUI interface that display the results of analysis for a single trajectory. The top left panel shows the calculated FRET trajectory; the middle two on the left are FRET autocorrelations and intensity correlations. The bottom panel on the left are the two intensity and sum intensity trajectory. The penals on the right are FRET histograms binned at 1 ms (top), user defined bintime (middle) and intensity histogram (bottom).

Figure 5.3. The SMS GUI 3 that displaying the individual trajectory information when incorporating it to the calculation of the ensemble value. The top left panel shows the calculated FRET trajectory; the middle two on the left are FRET autocorrelations and intensity correlations. The bottom panels on the left are the two intensity and sum intensity trajectory. The penals on the right are FRET histograms binned at 1 ms (top), user defined bintime (middle) and intensity histogram (bottom).
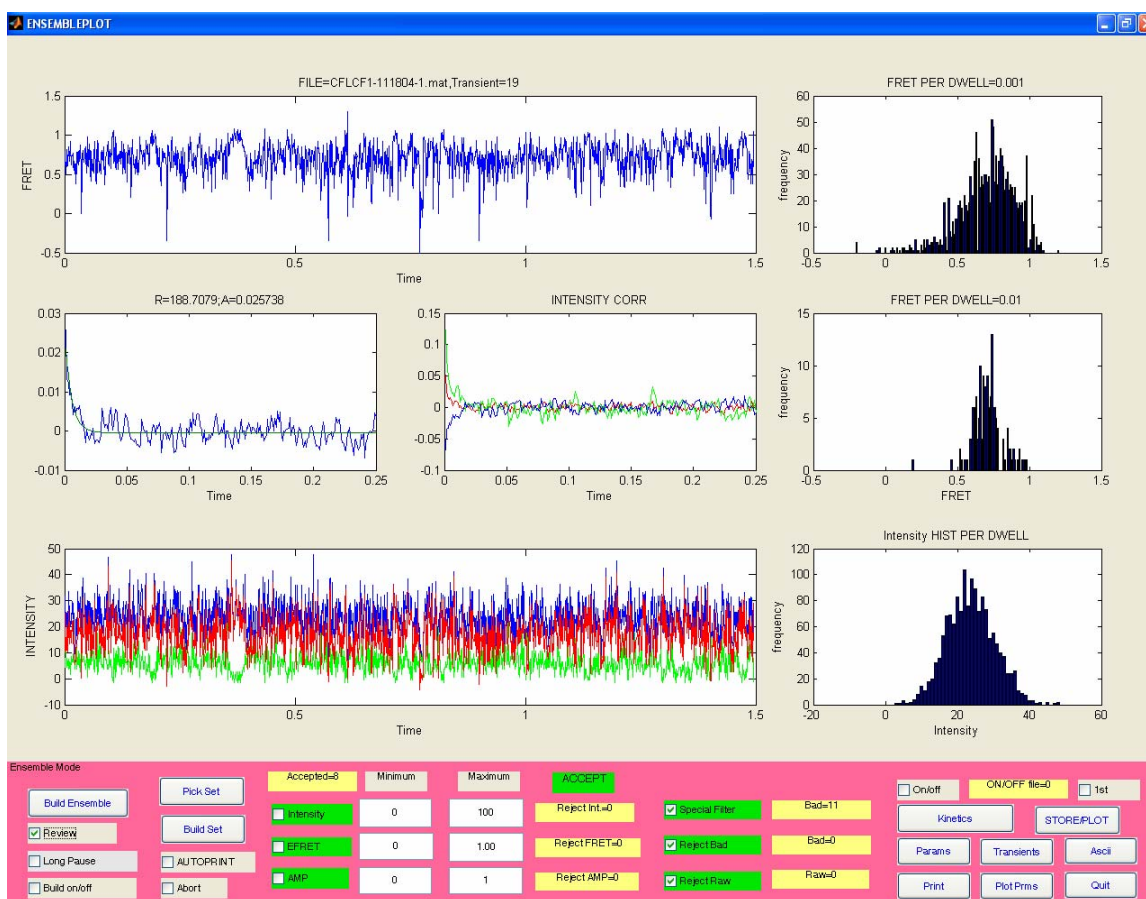
117
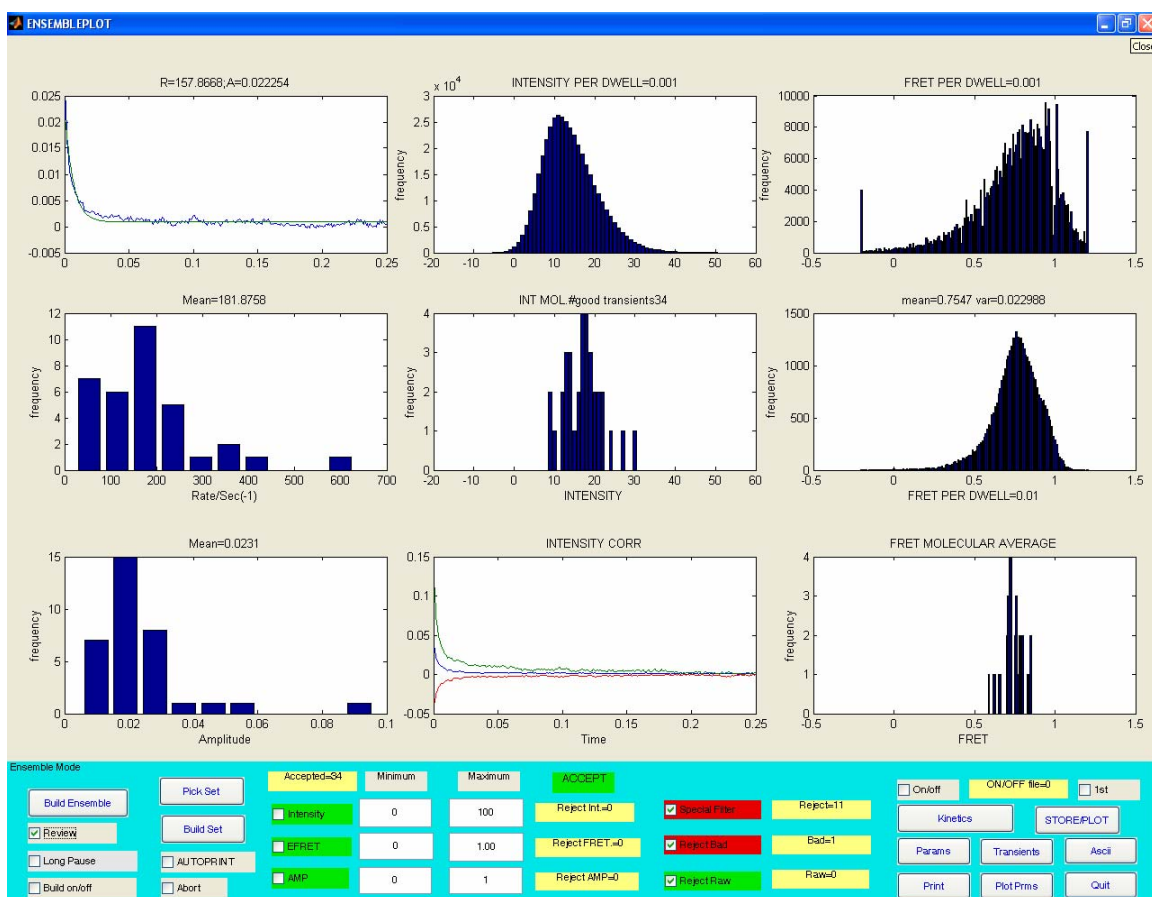
Figure 5.4 The SMS GUI 4 that displays the final ensemble results for a certain set of single molecule trajectories. The left panels are the FRET autocorrelation function fit with single exponential decay (top), the distribution of decay rate of the FRET autocorrelation function (middle) and amplitude of the FRET autocorrelation function. The middle panels are intensity distribution (top), the average intensity distribution (middle) and the average intensity autocorrelation and cross-correlation. The penals on the right are FRET histograms binned at 1 ms (top), user defined bintime (middle) and average FRET histogram (bottom).

Kinetic Mode of Scanning Confocal Microscope

The confocal configuration with APD detection is ideal for imaging thin samples at equilibrium and for collecting intensity transients with sub-millisecond and slower dynamics, which have been used in this dissertation to study single molecule intensity fluctuations due to FRET. Since the instrument uses single-spatial-point detection and excitation, and monitors only one molecule at a time, it is not well suited for irreversible reactions (which consume the sample after the reagents are mixed).

However, the scanning confocal configuration will be still sufficient for slow reaction kinetics as long as the reaction rate is so slow that the time scanning an image is much less than the reaction half-life time. In comparison to the trajectory-collection mode, we call this operation mode "kinetic mode". In the early configuration of the apparatus, the Digital Instruments stage PXY 100 is driven by *Nanoscope* V5.12 software. In the *Nanoscope* software, the images are saved, sometimes also flattened, in a special format. Therefore a Matlab routine has been developed to read the intensity information for those images (see Appendix I). This routine can also recover the data in those images flattened by *Nanoscope*. The stack of images will be analyzed by a Matlab routine called SMC to count the number of molecules in each image and draw a kinetic curve from a stack of images.

Normally, 7 uW excitation and 2 Hz scan rate are used, which corresponding to 120 seconds "dwell time" per image. For faster reactions, higher scan rate with higher excitation power can be used to increase time resolution.
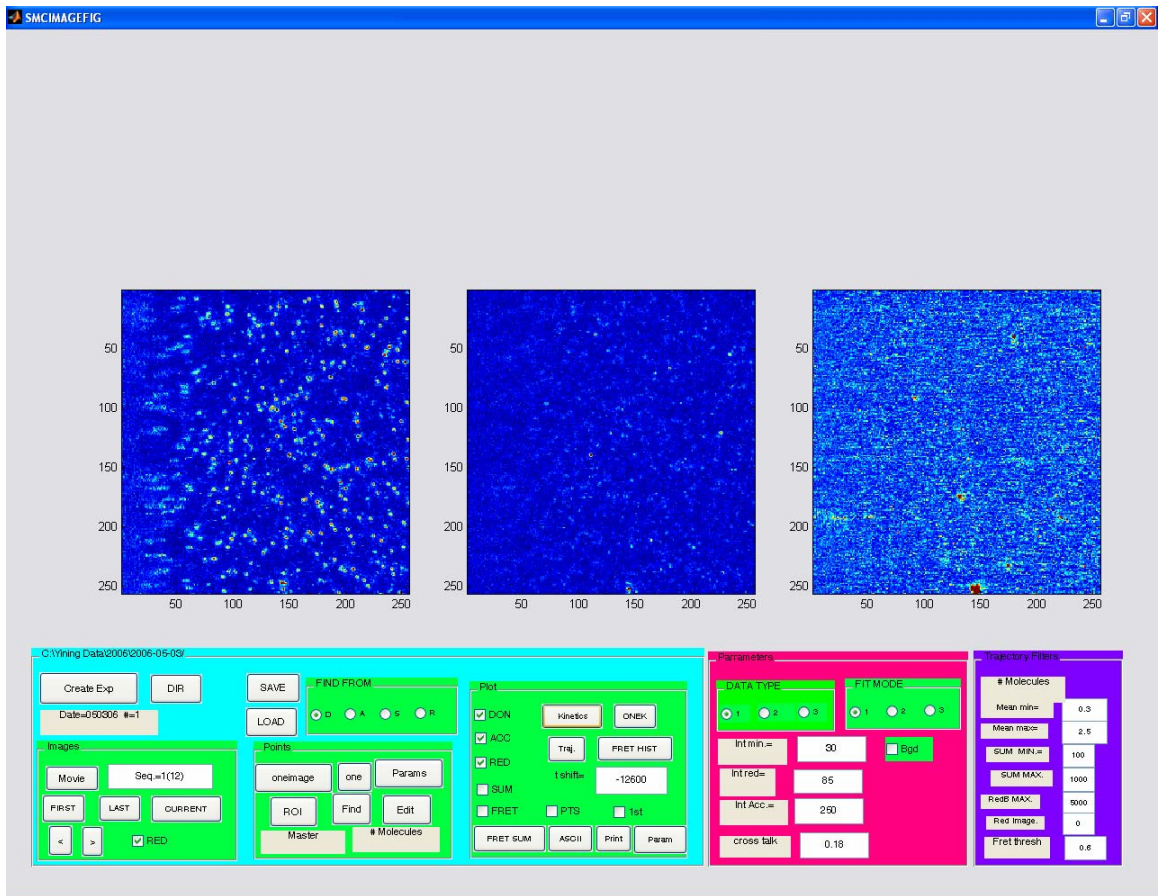
Figure 5.5 The SMC GUI that loads in images from scanning confocal microscope. This advanced version is also loading the red image for analysis.
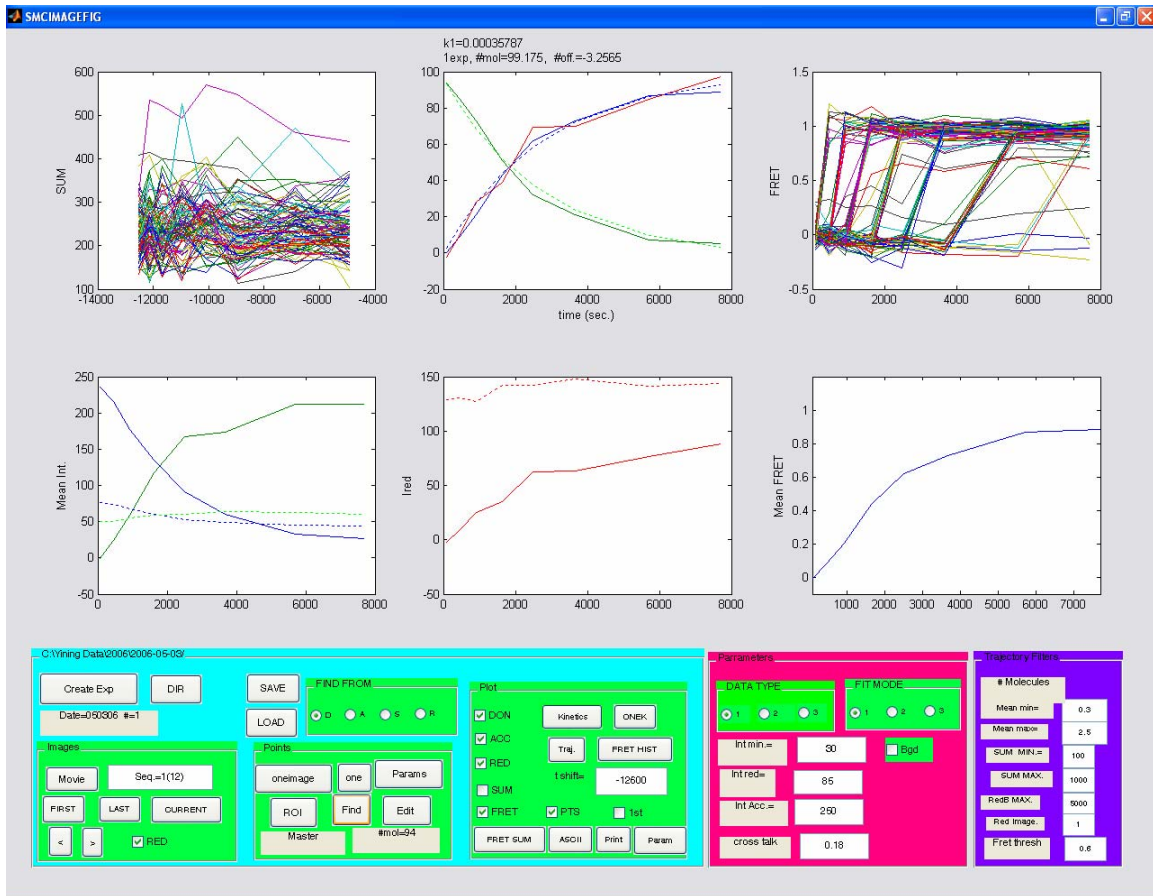
Figure 5.6 The sample SMC analysis of D-TAR and A-zipper annealing reaction. (Top) the left panel shows the intensity trajectories of a hundred molecules; the middle panel shows the number of donor, acceptor and red molecules; the right panel shows the FRET trajectories of those molecules. (Bottom) the left panel shows the donor/acceptor intensity and background intensity trajectories; the middle panel shows the red intensity and background intensity; the right panel shows the mean FRET value.

Figure 5.7 The statistic histogram information for one frame of images. (Top from left to right) single molecule red, donor and acceptor intensity. (Bottom from left to right) FRET histogram and red, donor and acceptor background intensity.

Two-Color Alternate Excitation Confocal Microscope

The single-color excitation microscope allows one detects donors and acceptors that are in close proximity of donors in order to have efficient FRET. However, when acceptors are too far away from donors to have efficient FRET, e.g. in enc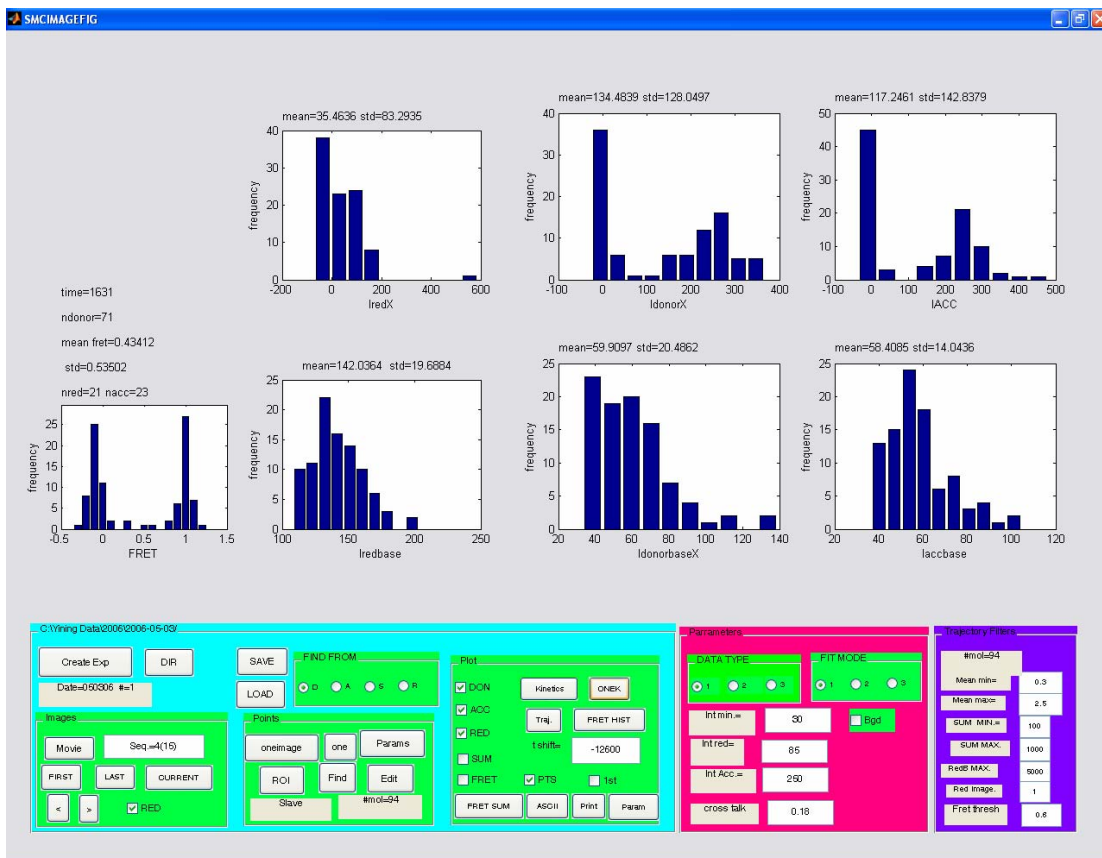ounter complexes and mis-annealed products, they are not longer visible under single-color excitation microscope. Under those circumstances, it is necessary to probe acceptors by direct exciting them with red laser. With this idea in mind, we build a microscope to simultaneously image donors and closely bound acceptors by exciting with green laser and image acceptors by exciting with red laser. The switching between two excitations is much faster than the scan rate, so that during the dwell time of each pixel, there are tens of switches. Therefore, each pixel has almost the same amount of time for green and red excitation. One scan will give fluorescence images for donor, acceptor (green excitation) and acceptor (red excitation), or red.

The on-off switching between two excitations is carried out by two ME-403 Acousto-Optic Modulators (AOM) produced by IntraAction Corp. Two AOMs are turned on alternatively by two 29 kHz square-wave signals, inverted-phase with respect to each other, generated from the same function generator. One is from the standard BNC connection port, the other is from the auxiliary BNC connection port, which is set as 180 degree phase shift with respect to the standard output signal. Considering that the AOM has a certain response time, there is a time lag set between the ON periods of the two square-waves by setting the symmetry to 40% instead of the ideal 50%.
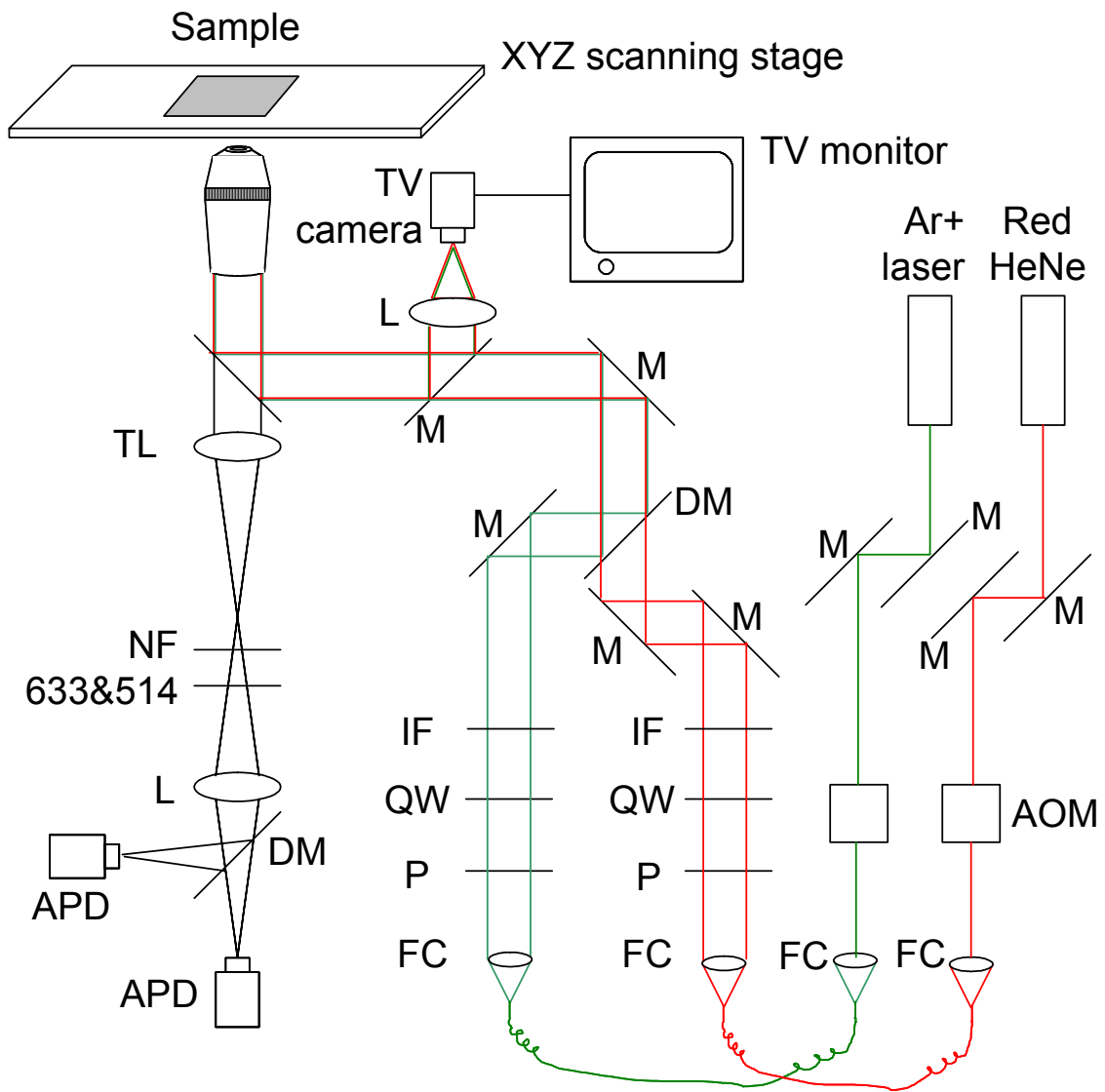
Figure 5.8 The scheme of two color alternate excitation microscope. AOM: acousto-optic modulator; APD: avalanche photo diode; DM: dichroic Mirror, FC: fiber coupler; IF: interference filter; L: lens; M: mirror; QW: quarter wave plate; P: polarizer; TL: tube lens.

Three counters on National Instruments PCI 6602 board are used to count donor, acceptor (green excitation) and acceptor (red excitation) or red intensities. The above two square-waves are used to gate the three counts. The square-wave signal for the green AOM is used to gate donor and acceptor counters and the other signal for red counts. In this way, donor and acceptor counter count intensity only when the green AOM is turned on, and the red counter counts intensity only when red laser is turned on.

I have developed a Labview software that drives the stage and scan images of the above three channels. The program is based on the single color-excitation Labview program generously provided by Prof. Douglas English at the University of Maryland. Several major changes have been made for two-color alternative excitation purpose.

First, three images from three independent counters have to be gated by outer signals. The 6602 counter board has eight counters. In Doug's Labview program, counter #0 and #1 are used for donor and acceptor intensity counting. However, I found that the Labview virtual instruments (VI) that are used for counter #0 and #1 are not compatible for the rest of the six counters. In my Labview program for two-color microscope, I used counter #3, #4 and #5 for donor, acceptor and red channels. These three counters are gated by the outer signals. The counter#2 still serves for timing the dwell time. The part of the VI that does the counter configuration is as follows:
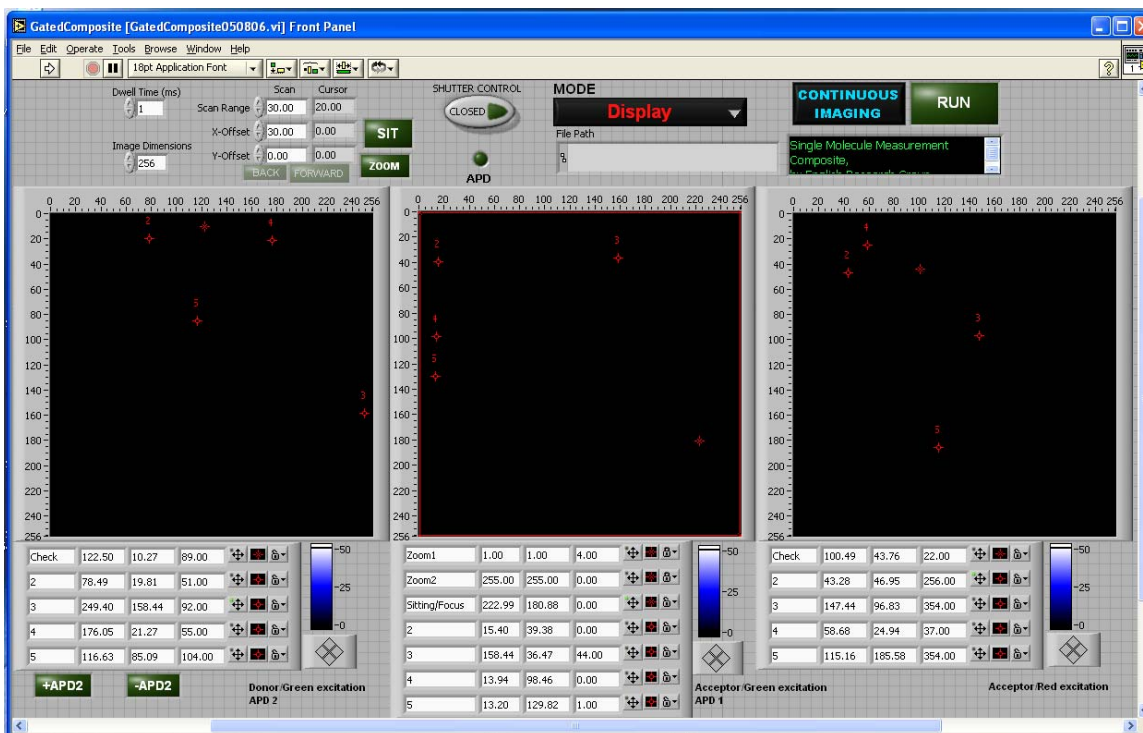
Figure 5.9 The front panel of the Labview for driving the two color excitation scanning confocal microscope.

The scan is carried out in a manner of line-by-line and pixel-by-pixel in each line. The dwell time of each pixel is defined by the input and used for the pulse width generated by counter#2, the counter that control the timing for each pixel. At the starting of each line, the three counters are reset. When the timing pulse is on, the stage moves to a new position, the counters will count intensities whenever the outer signal is on. When the timing pulse is off, the three counters' readout is exported, subtract the previous readout values to get the intensity for the current pixel. Then the stage moves to the next pixel in the same line. Once a line is finished, stage moves to the next line. Here is a chart flow of the scanning and counting process.

Because the counters are storing the total counts of intensity during each line scan, it is better to check if the counters might overflow during the scan. Each of the eight counters has the capacity of 32 bits, which are about 4294 million counts. When 254 pixels per line are used, each pixel has a capacity of 65 kilo counts. This value is well above the threshold set by our APDs, as discussed in below. Obviously, the counters are not saturated during the scan. However, one should be careful if more pixels are used when trying to obtain a fine image, as more pixels will imaginably consume substantial counting capacity.
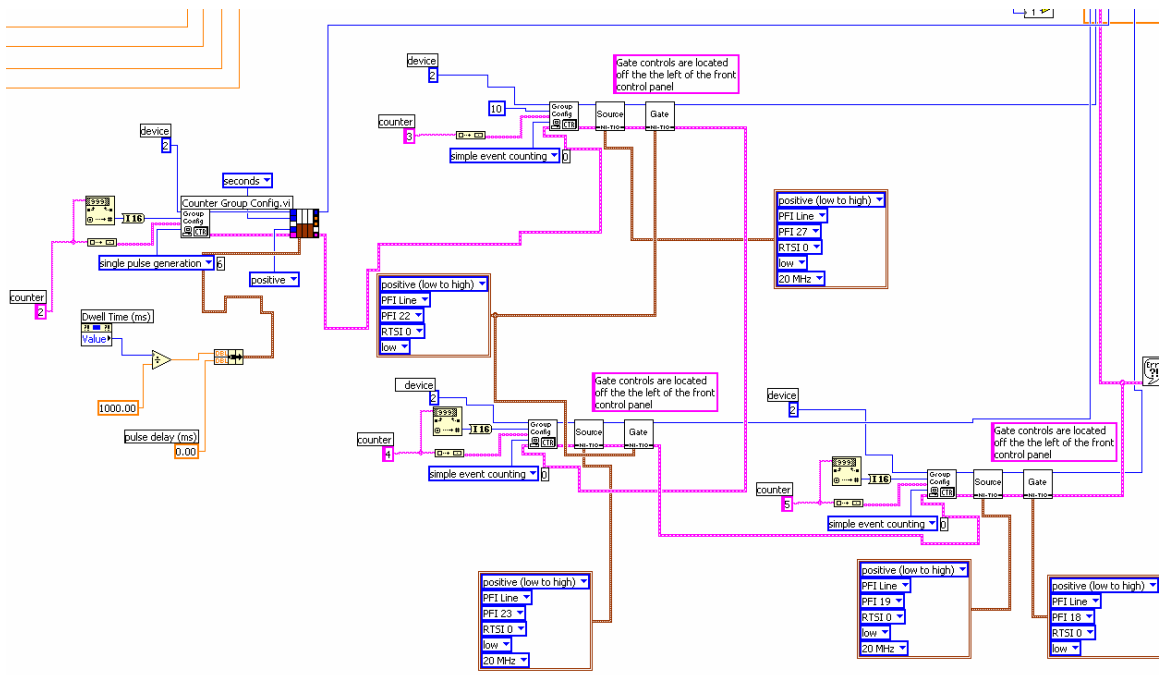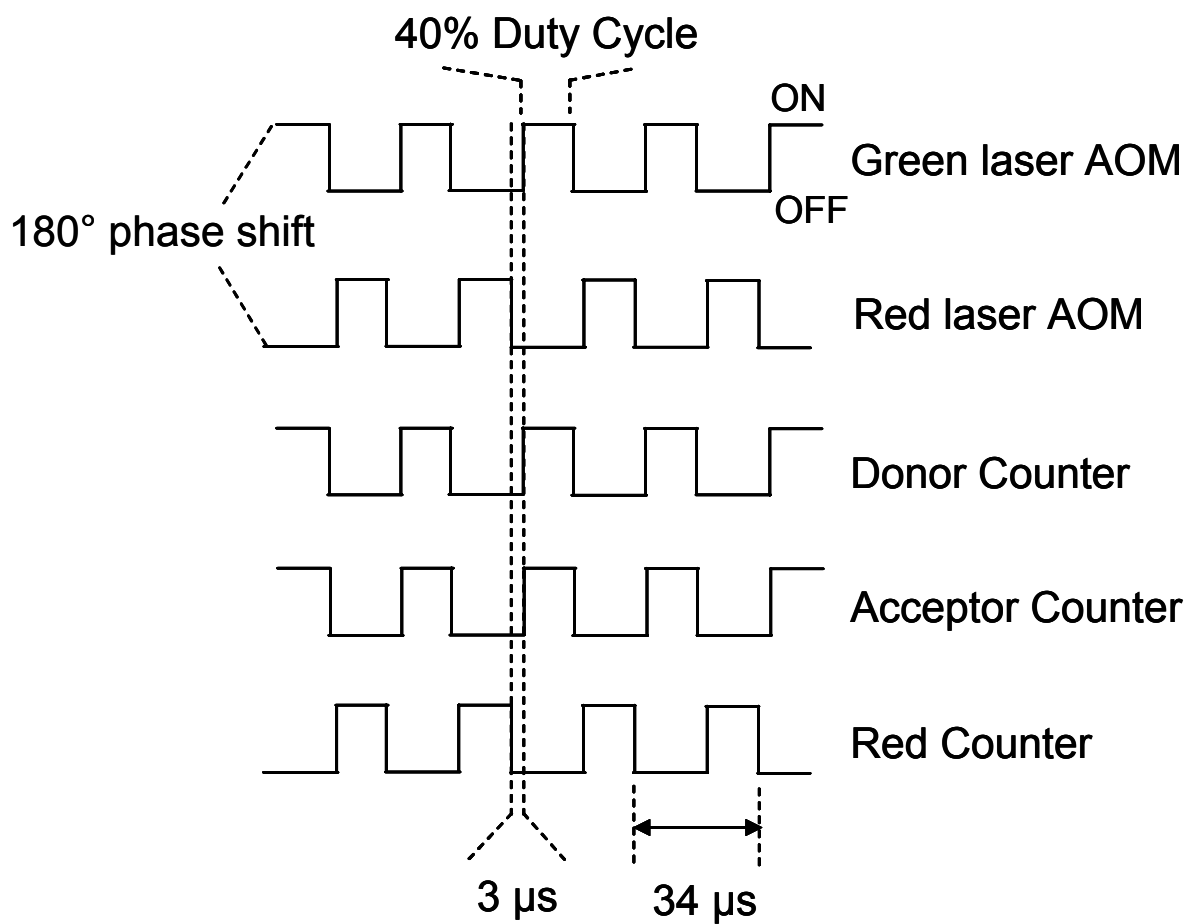
Figure 5.10 The Labview VI for counter configuration.

Figure 5.11 Pulse sequence for controlling of the AOM and counters.

```
┌─────────────────────────────────┐
│      Move one step in Y axis     │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Reset three counters, prepare to count │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────────────┐
│  Move one step in X axis and check its position │
└─────────────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│      Start all three counters    │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│          Wait for 1 ms           │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────────────┐
│  Read current total counts in the counters │
└─────────────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────────────┐
│  Calculate for intensity of one pixel in the line │
└─────────────────────────────────────────┘
```
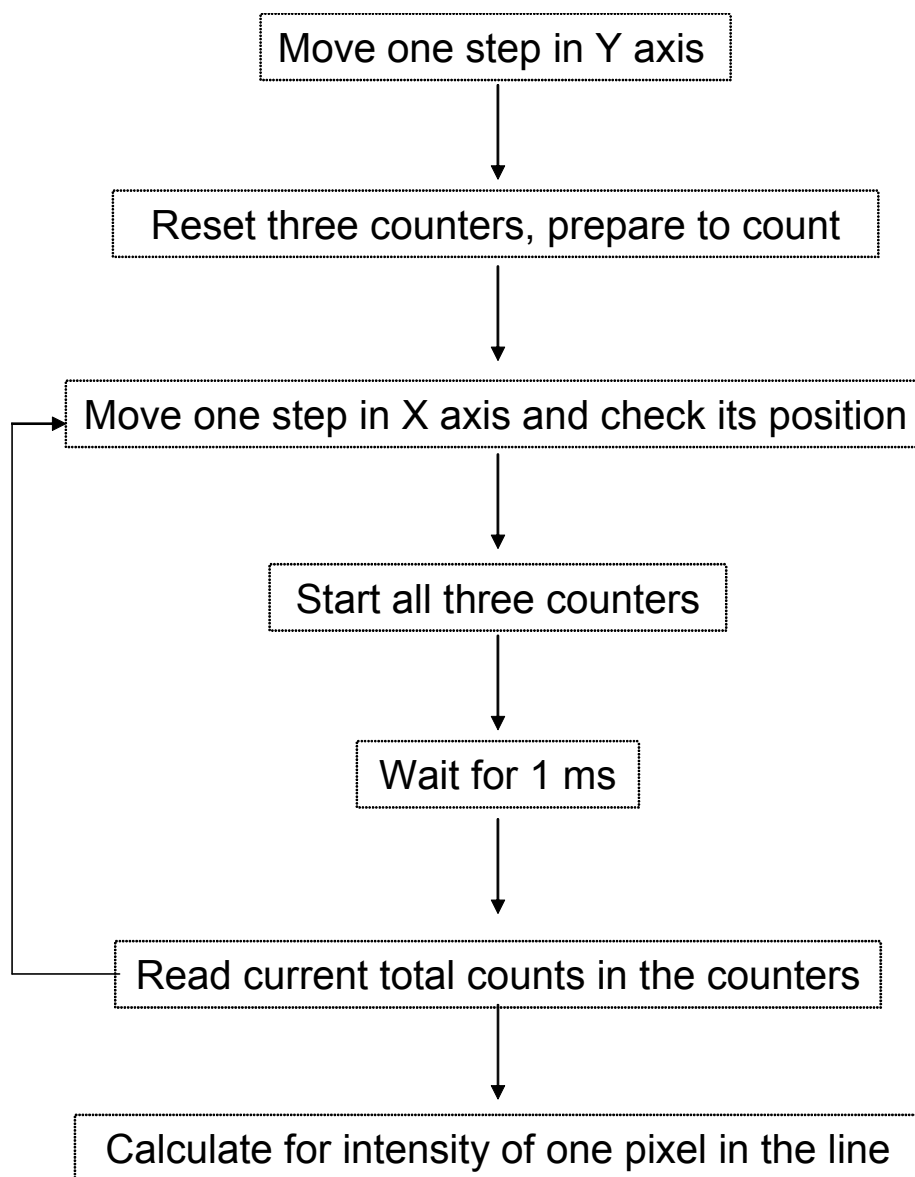
Figure 5.12 Chart flow for how the scanning and countering procedure.

The saturation of the detector is another important issue. In our microscope, we have two APDs as detectors; each has a photon counting capability of 1,000,000 per second. This value is equivalent to 3000 photon counts per pixel, considering that we have 3 ms dwell time for each pixel. In typical single molecule measurements, we normally have 50-100 counts/pixel. Even for beads sample used in the alignment of microscope, we have less than several hundred/pixel. Obviously, APDs will not be saturated in the experiment. The intensity data are saved as 16-bit integers, which has a capacity of 65536 counts per pixel, far beyond the limit that the experimental data can reach. So, the imaging program is not saturated, either.

The date and time information is saved to the beginning of each image file because the SMC program need them to establish kinetic data sets. Here is the part of VI that obtains the date and time information and saves it. All the images saved by Labview will be read by a Matlab routine into the SMC program, see appendix for the routine that reads Labview images. It is important to note that the Labview data should be read in as "Big Endian" format.

The back-reflection from the sample is collected and projected onto a video camera. The back-reflection spot is intentionally de-focused slightly. By monitoring the spot on TV camera, one can monitor the focus of microscope objective on the coverslip. The Z position of the scanning stage can be controlled by a user developed VI. During the experiments, the Z focus position is monitored on TV monitor and tuned by Lavbiew if necessary. In the future, it is possible to develop an automatic focusing mechanism based on the back-reflection spot.
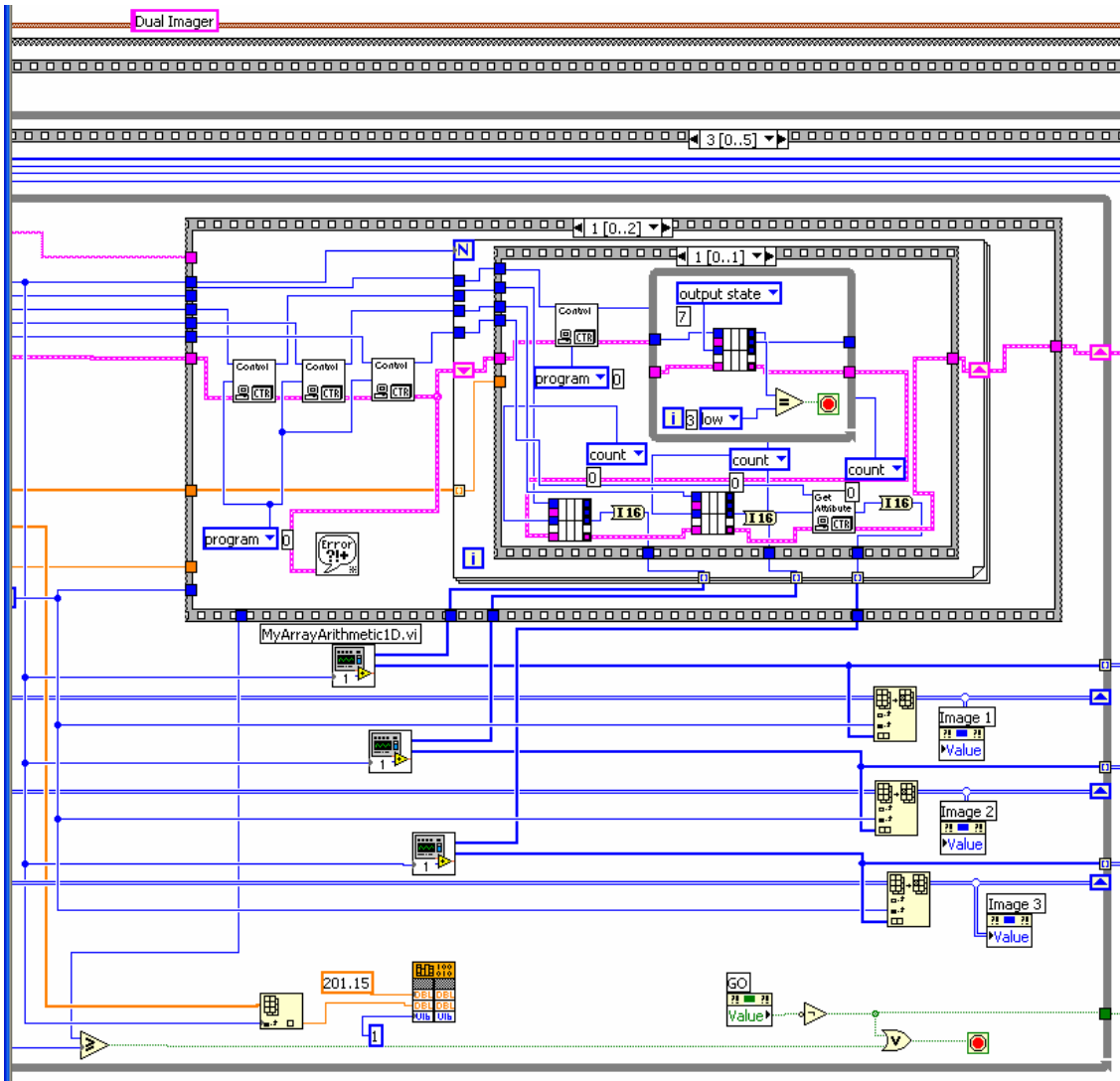
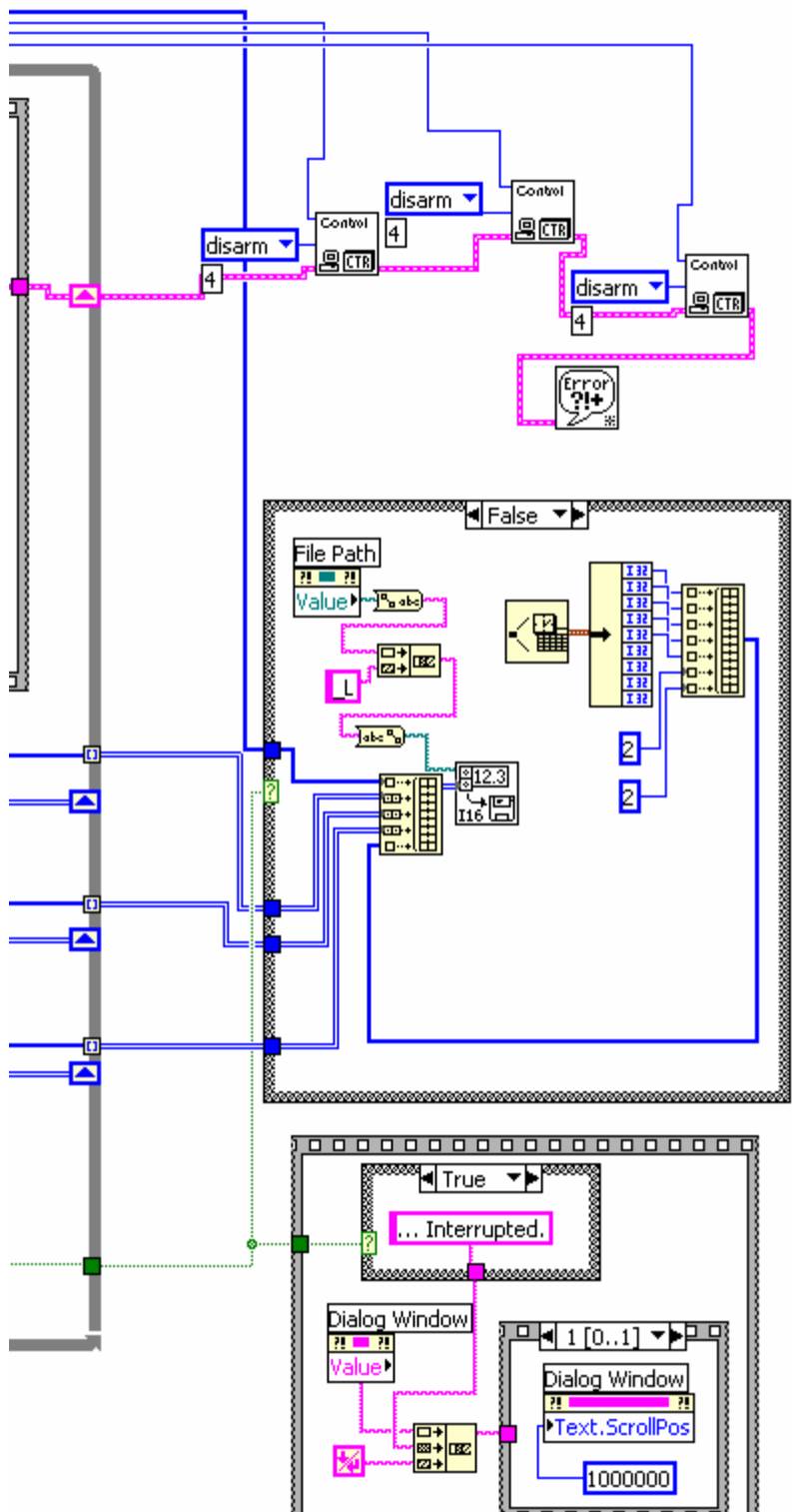Figure 5.13 The Labview VI to control scanning and counting.

Figure 5.14 The Labview VI to save the scanned three fluorescence images.
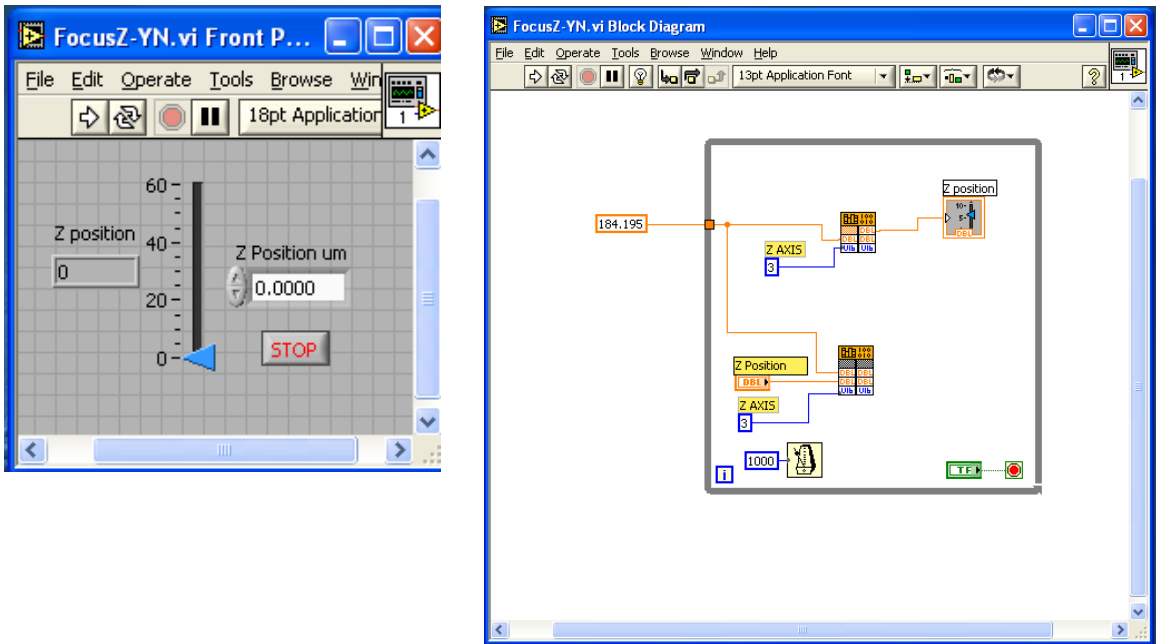
Figure 5.15 The Labview VI to control the Z focus position.

The Wide-Field TIR Microscope

The wide-field configuration is ideal for simultaneous monitoring of the fluorescence of many NC/nucleic acid complexes undergoing irreversible chemical reactions. Compared with the kinetic mode of confocal microscope, it has much better time resolution but lower signal to noise ration. However, more importantly, due to continuous excitation, dye photobleach become a serious compromise for the meansurement of slow kinetic reaction. Herein a Nikon wide-field microscope is introduced and its capability for single molecule annealing measurement are briefly discussed. The conclusion is that even it is sensitive enough for detecting large FRET change, this wide-field microscope has enough is not suitable for slow reactions unless proper chopping mechanism is involved to relieve photobleaching effect.

The Nikon TE2000-E wide-field TIR microscope uses through-objective total internal reflection (TIR) excitation at the upper surface of the bottom microscope cover slip. The TIR measurements utilize the exponential decay of the evanescent field generated on total internal reflection at a high-index to low-index interface, typically at the interface between coverlip and water solution[1,2]. When the incidence angle is above the critical angle, the evanescent field intensity (I) is proportional to the square of the electric field amplitude and levels off with distance (z) into the low-index phase.

$$I(z) = I(0)\exp(-z/d)$$

where I(0) is the evanescent field intensity at the interface. The distance factor in the exponential term is determined by:

$$d = \frac{\lambda}{2\pi}(n_2^2 \sin^2 \theta - n_1^2)^{-1/2}$$

where λ is the wavelength of excitation light in vacuum and n_1, n_2 are the indices of reflection of high-index and low-index medium, respectively. In our experiment, n_2= 1.518, n_1= 1.33, critical angle is around 61 degree and the d is around 130 nm.

Three laser lines, Ar 488 nm, HeNe 543 nm and HeNe 633 nm are equipped with microscope. In order to obtain higher excitation power for better signal-to-noise ration, a Millenia Ti-Sapphire laser that provides at least 200 mW (before microscope) 532 nm excitation is normally employed. The laser beam is focused onto the back focal plane of a microscope objective. For wide-field experiments, a 60x, N.A. 1.2, oil-immersion objective (PlanApo, Nikon) is used. The fluorescence is collected by the same oil-immersion microscope objective, separated into donor and acceptor channels, and then re-focused with 1.6x magnification factor to a Cascade 512B back-illumination CCD camera containing 512 by 512 pixel arrays. The field of view of the microscope is around 80 um, with each pixel on image equivalent to 160 nm.

The controlling software is Metamorph V 5.0. The measurements have been taken on 10 MHz port#1 with Gain set at 3 (which equals to 3.18 e.ADU) and on-chip-gain at 4090 (which equals to 1615x). No binning of pixels is performed. The exposure time is normally 50-500 ms.
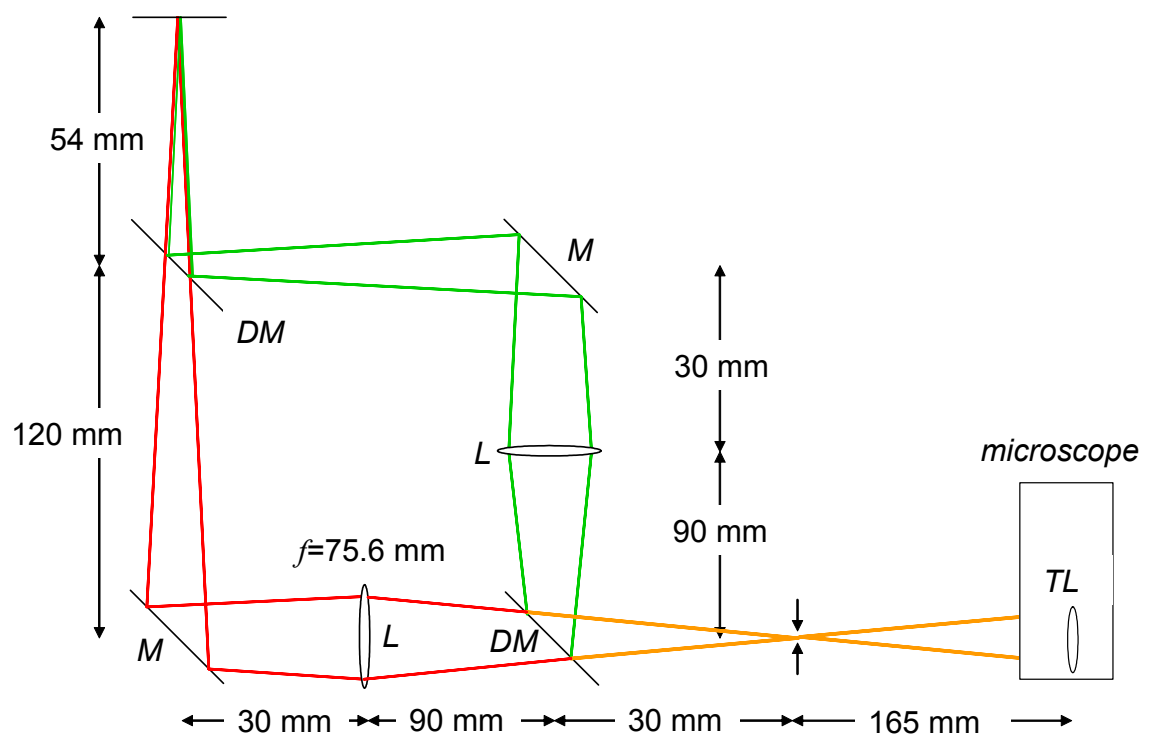
Figure 5.16 Optic scheme for detection of donor and acceptor single molecule fluorescence emission for TIR wide-field microscope. DM: dichroic mirror; L: lens; M: mirror; TL: tube lens.
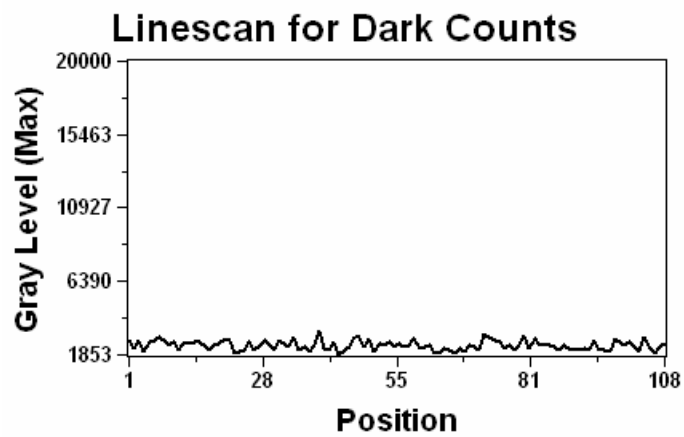
Figure 5.17 Dark counts measured by line scan. CCD settings are: 10 MHz port#1 with Gain 3 and on-chip-gain 4090, exposure time 100 ms.

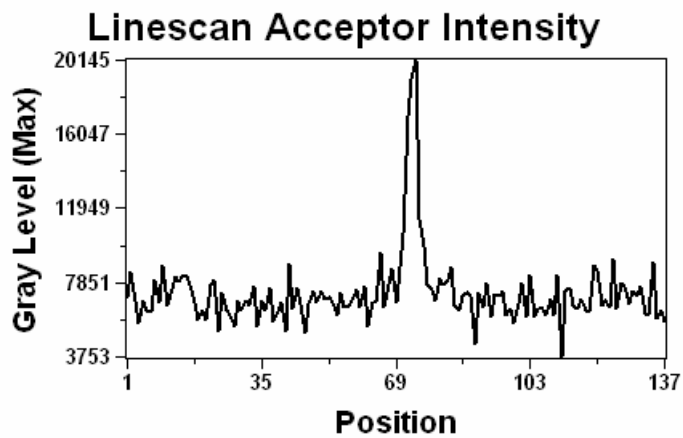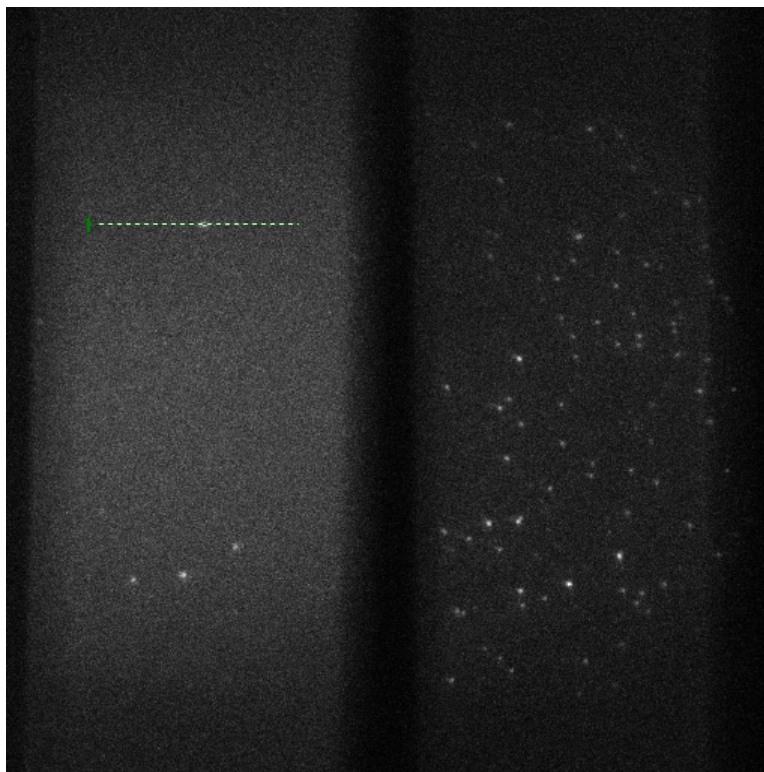Figure 5.18 Wide field Cy5 fluorescence signal-to-background ration measurements on Cy3, Cy5-labeled TAR DNA. CCD settings are: 10 MHz port#1 with Gain 3 and on-chip-gain 4090, exposure time 100 ms.
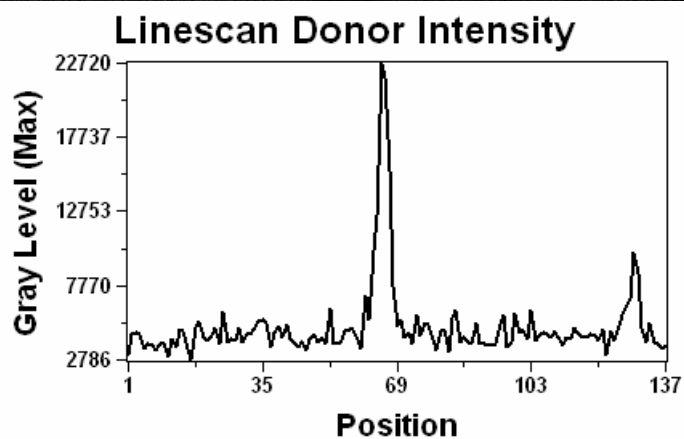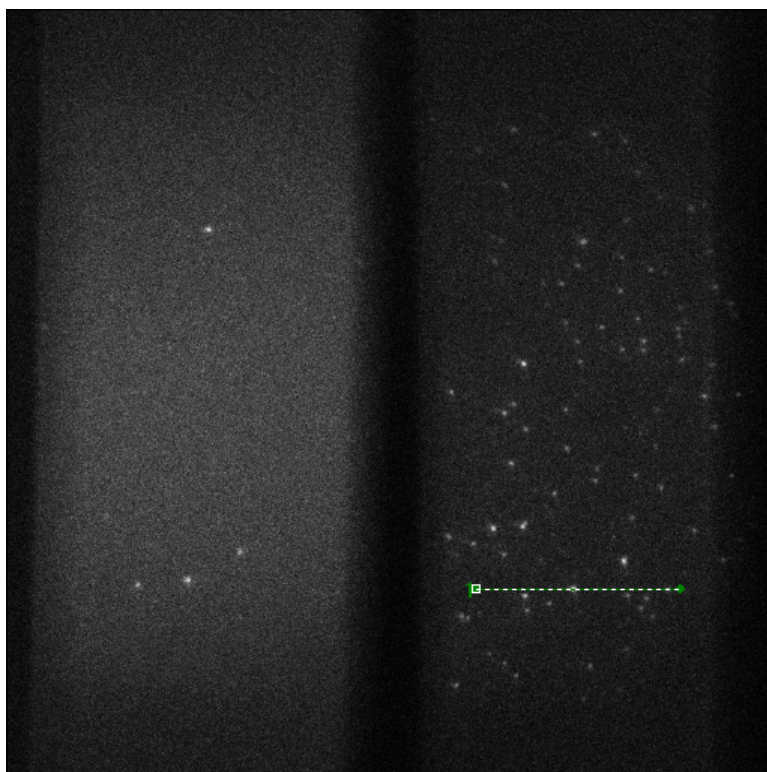
Figure 5.19 Wide field Cy3 fluorescence signal-to-background ration measurements on Cy3, Cy5-labeled TAR DNA. CCD settings are: 10 MHz port#1 with Gain 3 and on-chip-gain 4090, exposure time 100 ms.

However, due to its intrinsic limit, the signal-to-background ration (SBR) of wide-filed microscope is worse than that of confocal microscope, which is easily reaching 20 to 1. We have evaluated the SBR for the Cy3-Cy5 dye used for labeling. For Cy3, the SBR is typically around 6-10 to 1. The SBR of Cy5 is 4-5 to 1, which is a little bit worse. Our results are consistent to the literature reports for wide field single molecule SBR[3].

Because the donor and acceptor channels are independently focused onto CCD image plate, they have different positions and slightly magnification. Due to those factors, it is required to align the system with a fluorescence beads sample to determine the relative position of donor spot and acceptor spot on the image. An IDL program is generously provided by Dr. Rick Russell. First, one should manually specify the donor and acceptor pair. Three pairs are needed. Then the program will recognized the donor and acceptor position relationship in each pair and develop a linear equation to describe the position relationship. Then the program tries to find more pairs in the image and calibrate the equation. When the single molecule sample is measured, the program will used the position relationship determined from the beads sample to find the donor-acceptor pairs. Then intensity and FRET trajectory of each single molecule will be calculate by the program from the stage of images. Typically, trajectory for about 300 molecules can be obtained at 100 ms time resolution. The following is one particularly long-lasting single molecule trajectory obtained by wide-field microscope. The signal-to-noise ratio is good enough for detecting FRET change ~0.5 the 100 ms time resolution.
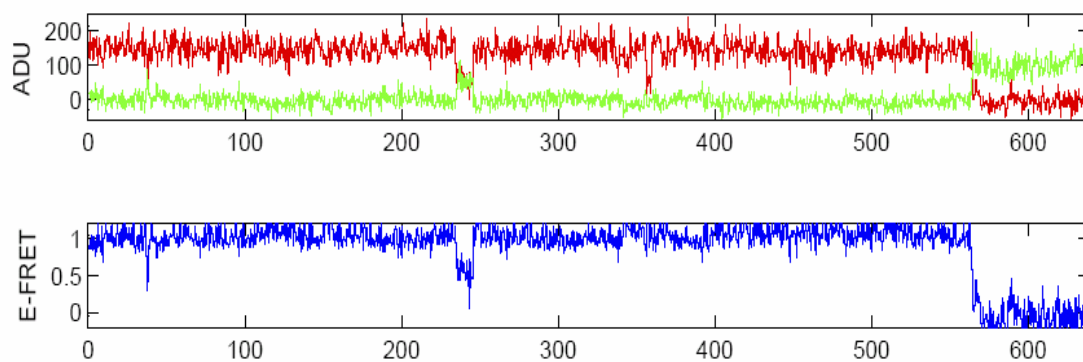
Figure 5.20 One of 297 intensity/FRET trajectories simultaneously captured by wide-field microscope obtained at 100 ms time resolution. The unit of horizontal axis is second. .

Appendix I The Matlab Routine That Recovers And Reads Images Digital Instruments Controlling Software

```
% Function to read tiff image files generated by DI confocal. Yining 05-11-2005
% Yining improve the reading speed, 03-16-2006
% the formats for input are as follows:
% date: 6 characters string
% serialnum: integer number
% filepath: the full/absolute path for the image
% the formats for the output are:
% Dat1 and Dat2 image matrix
% deltax and deltay: double
% time: character
function [Dat1,Dat2,deltax,deltay,time]=ReadImage(date,serialnum,filepath)
        if (serialnum>=0)&(serialnum<=9);
                SName=strcat('00',num2str(serialnum));
        end
        if (serialnum>=10)&(serialnum<=99);
                SName=strcat('0',num2str(serialnum));
        end
        if (serialnum>=100)&(serialnum<=999);
                SName=num2str(serialnum);
        end
   FName=strcat(filepath,date,'.',SName);
   FileID=fopen(FName,'r');
c(1,:)= fread(FileID, 10000, 'uint8=>char');
k=findstr(c,'\Scan size:');
fseek(FileID,k(1)+10,'bof');
   scansize(1,:) = fread(FileID, 7, 'uint8=>char');
   Scansize=str2double(scansize);
k=findstr(c,'\Scan rate:');
fseek(FileID,k(1)+10,'bof');
   scanrate(1,:) = fread(FileID, 8, 'uint8=>char');
   Scanrate=str2double(scanrate);
k=findstr(c,'\Date:');
fseek(FileID,k(1)+6,'bof');
   time(1,:) = fread(FileID, 8, 'uint8=>char');
k=findstr(c,'\Date:');
fseek(FileID,k(1)+15,'bof');
ampm(1,:) = fread(FileID, 2, 'uint8=>char');
        if
        (strcmp(ampm,'PM')&(str2num(time(1:2))~=12))|(strcmp(ampm,'AM')&(str2num
        (time(1:2))==12));
                Time=num2str(12+str2num(time(1:2)));
```

```
        time(1:2)=Time;
    end
kofs=findstr(c,'\Data offset:');
fseek(FileID,kofs(1)+13,'bof');
    offset1(1,:)=fread(FileID, 6, 'uint8=>char');
    Offset1=str2double(offset1);
fseek(FileID,kofs(2)+13,'bof');
    offset2(1,:)=fread(FileID, 6, 'uint8=>char');
    Offset2=str2double(offset2);
klen=findstr(c,'\Data length:');
fseek(FileID,klen(2)+13,'bof');
    datalen1(1,:)=fread(FileID, 7, 'uint8=>char');
    Datalen1=str2double(datalen1);
fseek(FileID,klen(3)+13,'bof');
    datalen2(1,:)=fread(FileID, 7, 'uint8=>char');
    Datalen2=str2double(datalen2);

ksline=findstr(c,'\Number of lines:');
fseek(FileID,ksline(1)+17,'bof');
  sampleline1(1,:)=fread(FileID, 3, 'uint8=>char');
  Sampleline1=str2double(sampleline1);
fseek(FileID,ksline(2)+17,'bof');
  sampleline2(1,:)=fread(FileID, 3, 'uint8=>char');
  Sampleline2=str2double(sampleline2);
fseek(FileID,Offset1,'bof');
    width1=Datalen1/(2*Sampleline1);
    Dat1=zeros(width1,Sampleline1);
for k=width1:-1:1;
    Dat1(k,:)=fread(FileID, Sampleline1, 'int16');
end
fseek(FileID,Offset2,'bof');
    width2=Datalen2/(2*Sampleline2);
    Dat2=zeros(width2,Sampleline2);
for k=width2:-1:1;
    Dat2(k,:)=fread(FileID, Sampleline2, 'int16');
end
kgg=findstr(c,'@Z magnify');
fseek(FileID,kgg(1)+24,'bof');
    zmag1(1,:)=fread(FileID, 11, 'uint8=>char');
    Zmag1=str2double(zmag1);
fseek(FileID,kgg(2)+24,'bof');
    zmag2(1,:)=fread(FileID, 11, 'uint8=>char');
    Zmag2=str2double(zmag2);
kgf=findstr(c,'MHz/LSB');
```

```
fseek(FileID,kgf(1)+7,'bof');
    zscale1(1,:)=fread(FileID, 10, 'uint8=>char');
    Zscale1=str2double(zscale1);
fseek(FileID,kgf(3)+7,'bof');
    zscale2(1,:)=fread(FileID, 10, 'uint8=>char');
Zscale2=str2double(zscale2);
kpf=findstr(c,'Plane fit');
kfd=findstr(c,'Frame direction');
fseek(FileID,kpf(1)+10,'bof');
    pfp1(1,:)=fread(FileID, (kfd(1)-kpf(1)-12), 'uint8=>char');
ps1=findstr(pfp1,' ');
fx1=pfp1(1:(ps1(1)-1));
pfx1=str2double(fx1);
fy1=pfp1((ps1(1)+1):(ps1(2)-1));
pfy1=str2double(fy1);
pofs1=pfp1((ps1(2)+1):(ps1(3)-1));
Poffset1=str2double(pofs1);
fseek(FileID,kpf(2)+10,'bof');
    pfp2(1,:)=fread(FileID, (kfd(2)-kpf(2)-12), 'uint8=>char');
ps2=findstr(pfp2,' ');
fx2=pfp2(1:(ps2(1)-1));
pfx2=str2double(fx2);
fy2=pfp2((ps2(1)+1):(ps2(2)-1));
pfy2=str2double(fy2);
pofs2=pfp2((ps2(2)+1):(ps2(3)-1));
Poffset2=str2double(pofs2);
planefit=pfp1((ps1(3)+1):length(pfp1));
Pfit=str2double(planefit);
klsba=findstr(c,'Sens. Counter');
klsbb=findstr(c,'MHz/LSB');
fseek(FileID,klsba(3)+17,'bof');
    sb1(1,:)=fread(FileID, (klsbb(1)-klsba(3)-19), 'uint8=>char');
lsb1=str2double(sb1);
fseek(FileID,klsba(5)+17,'bof');
    sb2(1,:)=fread(FileID, (klsbb(3)-klsba(5)-19), 'uint8=>char');
lsb2=str2double(sb2);
s=Sampleline1;
fclose(FileID);
        if Pfit==2;
                for x=1:256;
                        Dat1(:,x)=Dat1(:,x)+(-0.5*pfx1+(x-1)/(s-1)*pfx1)*lsb1*1000000;
                        Dat2(:,x)=Dat2(:,x)+(-0.5*pfx2+(x-1)/(s-1)*pfx2)*lsb2*1000000;
                end
                for y=1:256;
```

Dat1(y,:)=Dat1(y,:)+(-0.5*pfy1+(257-y-1)/(s-1)*pfy1)*lsb1*1000000;
Dat2(y,:)=Dat2(y,:)+(-0.5*pfy2+(257-y-1)/(s-1)*pfy2)*lsb2*1000000;
      end
    Dat1=Dat1+Poffset1;
    Dat2=Dat2+Poffset2;
  end
Dat1=Dat1.*(Zmag1*Zscale1*1000000)/65536+32768*(Zmag1*Zscale1*1000000)/65536;
Dat2=Dat2.*(Zmag2*Zscale2*1000000)/65536+32768*(Zmag2*Zscale2*1000000)/65536;
deltay=Scansize/Sampleline1;
deltax=deltay;
end

Appendix II Matlab Routines That Reads The Labview Images

```
% this function reads labview dual images to SMC program
% Yining 02-07-2006


function
[Ima1,Ima2,Ima3,Ima4,Imagetype,ExType,Dectype,ImaSize,time]=LVimage(date,serialnum,filepath)

lvdate=strcat('20',date(5),date(6),'-',date(1),date(2),'-',date(3),date(4));

    if (serialnum>=0)&(serialnum<10);
        fnA=strcat(filepath,'\',lvdate,'-II00',num2str(serialnum),'_L');
    elseif (serialnum>=10)&(serialnum<100);
        fnA=strcat(filepath,'\',lvdate,'-II0',num2str(serialnum),'_L');
    else
        fnA=strcat(filepath,'\',lvdate,'-II',num2str(serialnum),'_L');
    end

    [trcfid, message]=fopen(fnA,'r');
    fseek(trcfid, 4, 'bof');
    ImaSize=fread(trcfid,1,'uint16');
    if ImaSize<256;
        ImaSize=ImaSize+65535;
    end
    ImaSize=ImaSize/256;
```

```matlab
fseek(trcfid, 10, 'bof');
Imagetype=fread(trcfid,1,'uint16')/256;

fseek(trcfid, 2*ImaSize, 'bof');
switch Imagetype;
    case 1
        for k=1:ImaSize;
            Ima1(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
        end
        Ima2=0;
        Ima3=0;
        Ima4=0;
    case 2
        for k=1:ImaSize;
            Ima1(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
        end
        for k=1:ImaSize;
            Ima2(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
        end
        Ima3=0;
        Ima4=0;
    case 3
        for k=1:ImaSize;
            Ima1(k,:)= fread(trcfid,ImaSize,'int16','b');
        end
        for k=1:ImaSize;
            Ima2(k,:)= fread(trcfid,ImaSize,'int16','b');
        end
        Ima3=0;
        %for k=1:ImaSize;
        %    Ima3(k,:)= fread(trcfid,ImaSize,'int16','b');
        %end
        Ima4=0;
    case 4
        for k=1:ImaSize;
            Ima1(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
        end
        for k=1:ImaSize;
            Ima2(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
        end
        for k=1:ImaSize;
            Ima3(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
        end
        for k=1:ImaSize;
```

```matlab
                    Ima4(k,:)= fread(trcfid,ImaSize,'uint16', 'b');
                end
        otherwise
                error('This is impossible')
    end
    ss=fread(trcfid,1,'int16','b');
    mm=fread(trcfid,1,'int16','b');
    hh=fread(trcfid,1,'int16','b');
    time= [hh mm ss];
    dd=fread(trcfid,1,'int16','b');
    MM=fread(trcfid,1,'int16','b');
    ExType=fread(trcfid,1,'int16','b');
    Dectype=fread(trcfid,1,'int16','b');

    fclose(trcfid);


end
```

References

(1)   Daniel Axelrod EHHaRMF (2002) in *Topics in Fluorescence Spectroscopy* (Springer US, Vol. 3, pp. 289-343.

(2)   Axelrod D (1989) in *Total Internal Reflection Fluorescence Microscopy, in Methods in cell biology: Fluorescencemicroscopy of living cells in culture*, ed. D. TaYW (Academic Press, Inc., San Diego), pp. 245-270.

(3)   M. F. Paige EJB, W. E. Moerner (2001) *Single Molecule* 2: 191-201.

Vita

Yining Zeng was born in Changsha, Hunan Province, China on April 23, 1976, the son of Jijia Zeng and Zhenchun Nie. After completing his study at The High School Attached to Hunan Normal University, Changsha, Hunan, in 1995, he entered Peking University (or Beijing University) in Beijing, exempt from college admission exam due to his excellent academic record. He graduated with a Bachelor of Science degree in Chemistry in May 1999. He was award the Excellent Graduate of Beijing District upon graduate. During the summer of 1999, he was recommended and exempt from admission exam again to graduate school on Bio-Analytical Chemistry at Peking University with his advisor, Professor Yuanzong Li. He graduated with a Master of Science degree in Analytical Chemistry in May 2002. In August 2002, he entered the doctoral program in the Department of Chemistry and Biochemistry at the University of Texas at Austin with his advisor, Professor Paul F. Barbara.

Permanent address: 190 Shaoshan Middle Road, Changsha, Hunan, 410007, China

This dissertation was typed by Yining Zeng